



Title	Bringing about change in language education
Other Contributor(s)	University of Hong Kong. Dept. of Curriculum Studies.
Author(s)	Berry, Roger; Berry, Vivien; Nunan, David
Citation	
Issued Date	1995
URL	http://hdl.handle.net/10722/54844
Rights	Creative Commons: Attribution 3.0 Hong Kong License

BRINGING ABOUT CHANGE IN LANGUAGE EDUCATION

Edited by

**David Nunan
Roger Berry
Vivien Berry**

Department of Curriculum Studies
The University of Hong Kong

Bringing about change in language education

Published and distributed by
Department of Curriculum Studies
The University of Hong Kong
Pokfulam Road
Hong Kong

ISBN: 962-8109-01-4

© Department of Curriculum Studies

Copyright

All rights reserved. No parts of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of the publisher.

International Language in Education Conference is organised by:
Department of Curriculum Studies, University of Hong Kong
The English Centre, University of Hong Kong
The Hong Kong Institute of Education
Hong Kong Association for Applied Linguistics

The conference organisers wish to express their gratitude to the **Language Fund** whose generosity has enabled these selected papers to be published.

Editing and layout: Allen Wan, The English Centre
Cover design: Vincent Lee, Department of Curriculum Studies

To John Harris

Contents

Introduction

- David Nunan, Roger Berry and Vivien Berry i

Chapter 1

- The more things change, the more they stay the same: Or why
action research doesn't work 1
— David Nunan

Chapter 2

- ✓ Helping teachers change the language of the classroom: Lessons
from in-service teacher education 21
— Philip Hoare & Stella Kong

✓Chapter 3

- Assessment and change in the classroom. 31
— Vivien Berry, Peter Falvey, David Nunan, Martin
Burnett & John Hunt

Chapter 4

- The examination-classroom backwash cycle: some historical
cases 55
— Bernard Spolsky

Chapter 5

- Washback or washout? The relationship between examination
reform and curriculum innovation 67
— Stephen Andrews

Chapter 6

- Methodology washback - an insider's view 83
— Lam Hiu Por

Chapter 7

- Towards a broader view of authenticity in the language
classroom: a discussion paper 103
— Patricia Galien & Wendy L. Bowcher

Chapter 8	
Reading and problem solving: the interaction of individual differences and task types	121
— Peter Storey	
Chapter 9	
"Willingness to communicate" among Hong Kong students	149
— Barry Asker	
Chapter 10	
Implementing quality in ESP training: experiences from the Hong Kong Vocational English programme	161
— Grahame T. Bilbow	
Chapter 11	
Assessing corporate foreign language needs: creating a language auditing toolkit adequate to meet international and cross-cultural needs	171
— Nigel Reeves	
Chapter 12	
Mother tongue education in a multilingual, multicultural society: the Malaysian Chinese experience in the 1990s	183
— Ng Kwei Kuen	
Chapter 13	
The quantum leap: from essay to thesis	201
— Linda Cooley & Jo Lewkowicz	
Chapter 14	
Dialogue journals and course evaluation	217
— Jonathon Shaw & Terry Clayton	

Introduction

The chapters in this volume were selected from a larger collection of papers originally presented at the International Language in Education Conference held at the University of Hong Kong, 14-16 December 1994. A major thematic concern draws the chapters together and helps to define them as a coherent whole. This is the notion of *change* in language education. At present, within Hong Kong, and more generally throughout the Asia Pacific region, there are major changes proposed within the school and tertiary educational sectors. This imperative for change is being driven, among other things, by emerging economic realities, the development of competency based education, and the globalisation of communication resulting from the explosion in the use of the Internet.

It is no coincidence that a majority of the papers in this collection deal with assessment in one form or another. Anyone with more than a passing knowledge of curriculum development knows that those who control the means of assessment also control the means of advancing or retarding educational change. It is hardly surprising then that terms such as 'change', 'assessment', 'washback', and 'evaluation' punctuate the chapter titles and papers in this collection.

It is also no coincidence that in these chapters explicit links are drawn between the world of education and the world of work. Due to the revolution currently taking place within the workforce, employers want school and university graduates to be creative, innovative, flexible, to be able to work efficiently in small teams, and, above all, to be able to communicate effectively. Governments and ministries of education throughout the region are getting the message that traditional methods of teaching and learning are not the most effective ways of producing graduates with these qualities.

The themes of the chapters in this volume illustrate these fast emerging realities. Major themes include the following:

- the use of action research as an instrument of change
- factors implicated in helping teachers change the medium of classroom instruction;
- the implications for assessment within the Target Oriented Curriculum initiative
- the role of the learner in the assessment process
- the language proficiency of Hong Kong graduates
- examination washback and curriculum change
- authenticity and the language classroom
- individual learner differences and test taking behaviour

- innovation and change in vocational training
- changing communication needs in the corporate world
- mother tongue education in a multilingual society
- the writing needs of post-graduate research students
- the use of dialogue journals in course evaluation

The value of this collection is that it provides theoretical and empirical insights into the dynamic and interactive forces that are having a mutual effect upon the classroom and the workplace within Hong Kong and beyond. The interested reader will also find many practical suggestions for creative and educationally justifiable ways of dealing with these pressures for change.

The Editors

Hong Kong, October 1995

1

THE MORE THINGS CHANGE THE MORE THEY STAY THE SAME: OR WHY ACTION RESEARCH DOESN'T WORK

David Nunan

English Centre, University of Hong Kong

"Can the quest for objectivity distract us from the search for truth?"
(Walker 1982: 57)

Introduction

In this paper, I shall explore some of the problematic aspects of attempting to bring about change in language classrooms through action research. Data for the study will take the form of case studies of action research projects carried out in a variety of EFL and ESL settings. Each of these projects encountered serious problems which jeopardised their very existence. In each instance, I shall outline the evolution of the project, set out what went wrong, and describe what was done in an attempt to remedy the problems. The paper will conclude with a set of operating principles for bringing about change in the classroom through the development of an action research agenda.

What Is Action Research?

Action research involves systematically changing some aspect of one's professional practice, usually in response to some issue, problem or puzzle, collecting relevant data on the effects of the changed practice, and the interpreting and analysing that data. In educational contexts, these procedures are meant to improve the quality of learning, and to enhance the rationality and justice of the educational system (Kemmis and Carr, 1986).

Numerous reasons have been advanced for the use of action research in professional development. Wallace, for example, argues that (1991:56-7), action research can be attractive for two particular reasons:

1. It can have a specific and immediate outcome which can be directly related to practice in the teacher's own context.
2. The 'findings' of such research might be primarily specific, i.e. it is not claimed that they are necessarily of general application, and therefore

the methods might be more free-ranging than those of conventional research.

'Research' of this kind is simply an extension of the normal reflective practice by many teachers, but it is slightly more rigorous and might conceivably lead to more effective outcomes." Those of us who work in teacher education know that one of the most difficult things to balance in a course is the tension between theoretical and practical aspects of the profession.

van Lier, (1992: 3) also argues persuasively for the use of action research as a way to bridge the gap between theory and practice, stating that:

... theory and practice are not perceived as integral parts of a teacher's practical professional life. ... This situation is the result of communication gaps caused by an increasingly opaque research technocracy, restrictive practices in educational institutions and bureaucracies (e.g. not validating research time, or not granting sabbaticals to teachers for professional renovation), and overburdening teachers who cannot conceive of ways of theorizing and researching that come out of daily work and facilitate that daily work. (van Lier 1992: 3)

Finally, Allwright and Bailey, (1991), have pointed to the value of classroom observation, reflective teaching and action research

Slowly, the profession as a whole is realising that, no matter how much intellectual energy is put into the invention of new methods (or of new approaches to syllabus design, and so on), what really matters is what happens when teachers and learners get together in the classroom. This shift in emphasis from concentrating on planning decisions to concentrating on looking at what actually happens in the classroom, has led researchers to have much greater respect for classroom teaching. The more we look, the more we find, and the more we realise how complex the teacher's job is. And teachers, in their turn, faced at last with researchers who have at least some idea of the enormous complexity of everyday classroom life, are beginning to be more receptive to the whole research enterprise. Being a good classroom teacher means being alive to what goes on in the classroom, alive to the problems of sorting out what matters, moment by moment, from what does not. And that is what

classroom research is all about: gaining a better understanding of what good teachers (and learners) do instinctively as a matter of course, so that ultimately all can benefit. (Allwright and Bailey, 1991).

Action research typically evolves through a series of cycles beginning with the identification of a puzzle or problem, some sort of preliminary investigation to obtain baseline data on the issue in question, the formulation of an intervention strategy, activation of the strategy and documentation of the results, reporting on the outcome, and planning the next cycle in the process. These steps are illustrated in Table 1 and 2, below. Table 1 summarises a study taken from a second language context, and Table 2, a study from a foreign language context.

Steps in the Action Research Process

Table 1: The Action Research Cycle: An ESL Example

1 Problem/puzzle Identification	--->	A teacher identifies a problem/puzzle. "My students don't seem interested or motivated."
2 Preliminary Investigation	--->	What's going on? Recording and observing class over several days.
3 Hypothesis	--->	Content doesn't seem to stimulate students. Exclusive use of display questions.
4 Plan intervention	--->	Increase use of referential questions. Make links between content and learners.
5 Outcome	--->	More complex interactions. More involvement and interest. More 'natural' discourse, e.g. students nominate topics, Ss disagree with teacher, S-S interaction
6 Reporting	--->	Staff development session

Table 2: The Action Research Cycle: A Foreign Language Example

1 Problem Identification	---	A teacher identifies a problem in her classroom "My students aren't using the target language"
2 Preliminary Investigation	---	What's going on? Recording and observing class over several days.
3 Hypothesis	---	Teacher uses too much English. The important stuff is done in English
4 Plan intervention	---	Teacher increases target language use. Teacher uses German for classroom management etc.
5 Outcome	---	Increase in use of German by students
6 Reporting	---	-Article in teachers' newsletter

Case Studies

Case Study 1; Phil the Enthusiast

The Context

In the 1980s, I was approached by the coordinator of a programme for providing professional support to foreign language teachers in Australia. The coordinator, let's call him Phil, had heard me give a talk on ways of empowering classroom teachers by handing power and responsibility back to them for critical aspects of their work such curriculum planning and development through school based curriculum development, and professional development, through classroom observation and action research. Phil wanted to know whether I was interested in working with him to plan, implement and evaluate a programme of professional development and renewal for a group of secondary school teachers of a wide range of foreign languages, including Indonesian, French, German, Chinese, Italian and Vietnamese. Excited at the prospect of systematically implementing my ideas with a different group of teachers from the ones for whom the ideas had originally been developed, I agreed on two conditions: firstly, that the project be over a reasonable span of time, that is two to three years, and secondly, that the Ministry of Education should provide the teachers taking part in the project with eight full days of paid release time during the course of the project. The first condition was readily agreed to. The second involved considerable negotiation. The final compromise was that the Ministry would pay for eight half-days, on condition that teachers contributed some of their own time. A classic case of Indian Giving, but it did result in the project getting under way.

The stimulus for change

Apart from the enthusiasm of Phil, as the project coordinator, and me as the project consultant, what was the spark that had ignited what was merely a good idea at the time? The single most important stimulus for the project was the imminent appearance of a highly controversial set of curriculum guidelines for the teaching of all languages other than English in Australian schools. Called appropriately enough the ALL Guidelines, these proposals, to be properly implemented, would require a major shift in the way that many foreign language teachers conducted their daily professional lives. The project that I became involved in was designed as a consciousness-raising exercise to encourage teachers to reflect on their teaching, to collect objective data on what went on there, and to identify areas where they wanted to improve.

What happened

The project succeeded with some teachers who felt that, despite the anxiety and amount of work involved, they had developed a more reflective attitude towards their teaching, they had identified and solved some of the problems with which they were confronted in their daily lives, and they had a greater sense of control over their own professional destiny. For some of the other teachers, the result was not quite so happy. They felt that they needed support on a day-to-day basis, rather than the half-day consultations that the project provided for. Some felt devalued and alienated within their schools (although others reported that their status had been enhanced by involvement in the project). Many were unable to conceptualise and focus their investigation, and this led to feelings of frustration and even anger. Finally, and inevitably, there were those who were fearful of being observed.

Emboldened by the general success of our initial efforts, Phil and I decided on an expanded project in the second year. This time the results were not quite so happy. The Government, as a condition of funding the project, insisted that the action research projects be explicitly tied to the implementation of a new set of curriculum guidelines. In effect, the government was saying that they would fund the research only if it gave them answers to questions arising from their curricular innovation. They also reduced the amount of time made available to teachers outside of the classroom to obtain expert help, to collaborate with colleagues, and to report on their work. Several teachers, believing that they were to become pawns of the bureaucracy, refused to continue their association with the project.

The stricture that teachers could only research an issue or area of direct interest to the funding authority violates one of the fundamental principles of action research - the freedom of the individual concerned to determine what to

investigate, how to investigate, and how to report. This particular initiative underlines the importance of adhering to fundamental principles.

Case Study 2: Marion the Martyr

The context

The second case study concerns a group of foreign language teachers working at the secondary school level. All teachers were non-native speakers of the language they were teaching. While most had trained as teachers of European languages, typically German and French, they were in the process of changing to the teaching of newly popular Asian languages such as Japanese, Indonesian and Chinese.

The stimulus for change

The stimulus for change in this case was a government mandated decree that all secondary students, and eventually all primary level students should study a second language. A dilemma was created for the educational system when students voted strongly with their feet to take Asian rather European languages. The dilemma was brought about because resources within the system were geared towards European rather than Asian foreign languages. The main problem was the paucity of teachers for Asian languages, particularly Japanese. As a result, pressure was placed on teachers to switch languages, and undergo crash courses in their chosen language as well as undergoing retraining in classroom methods and techniques.

One of the major planks in the professional development programme was the initiation of an action research project. it was felt that such a project would not only assist the teachers in redirecting skills they already possessed from the teaching of European to Asian languages, but that it would also connect with the real day-to-day problems and challenges that teachers were facing.

What happened

The teachers were given an initial two-day, paid release period to attend a workshop designed to familiarise them with the principles of action research, and with techniques for collecting and analysing data. They were also given three half-day follow-up sessions to work with other participants in the programme as well as consulting with senior advisers and the project consultant.

The problems began once teachers had taken part in an initial classroom observation and data collection exercise. All teachers were horrified at their level of competence in the target language, as revealed by the audio and video

recordings, and most were also depressed by the fact that their professional skills also suffered. For several the work involved in classroom observation and action research added an intolerable burden to what was already an extremely stressful period in their professional lives. Most reported that having to juggle and monitor recording equipment seriously interfered with their teaching duties, and the demands of conceptualizing their project (for example, focusing the question so that it was actually capable of being answered) compounded the anxiety and guilt that many felt at being, as one teacher put it, "only one lesson ahead of my students".

Case Study 3: Colin the Cynic

The context

The third case study took place in a foreign language context. The language institute concerned is devoted primarily to the teaching of English as a foreign language to post-secondary school students. There is also a vocational dimension to the course of studies. In general, the students entering the institution lack motivation, and consider that they have failed the system, their parents, and themselves.

The stimulus for change

Falling enrollments, brought about by changing demographics and also the prolonged recession of the 1990s led the school to reevaluate its place in the educational market-place, its curriculum, and its general approach to education. It decided that a major shift in focus was needed, and established a Curriculum Renewal Project to revamp the organisational structure of the institution, reconceptualise the curriculum following principles of learner-centredness and task-based learning, produce materials in-house to support the curriculum, develop an assessment system consistent with the new curriculum, and implement a system of professional development.

What happened

Overall, the project, which is ongoing, is a considerable success. This is particularly so in relation to the development of curriculum guidelines, the creation of materials, and, surprisingly, the development of an assessment system. The latter is surprising because, like all other aspects of the curriculum, it was developed in-house, by teachers who, initially at least, had little experience in the areas of testing and assessment.

The one area where the project ran into trouble was in the area officially designated as "professional development". I say "officially" designated, because, in actual fact, a massive amount of professional development

occurred as teachers became involved in activities associated with the renewal project such as developing syllabuses, assisting in the organisation of a self-access centre, writing materials, and developing assessment instruments and procedures. While there were several initiatives in the professional development area, including staff seminars, the centrepiece was intended to be a series of action research projects. It was felt that action research would have two major benefits. In the first place, it would give teachers control over their own professional development by enabling them to decide what they wanted to change in relation to their own teaching practice and how they wanted to change, and that it would also connect professional development with real issues, problems, puzzles and challenges. (All too often, formal seminar programmes, around which most professional development is organised, while interesting in their own right, fail to provide participants with solutions to problems they are having in their classrooms).

Despite the fact that the action research initiative was enthusiastically supported by the curriculum renewal consultants and teachers were given a practical introduction to techniques for identifying and refining questions, collecting and analysing data, and writing up the project, the action research initiative, after a faltering start, died completely.

Several things conspired to bring the action research initiative to a halt. In the first place, the administration decreed that no time off would be allowed for participants to carry out essential tasks such as transcribing data, consulting with colleagues, and writing up their results. Secondly, the teacher who was to coordinate the initiative worked to undermine it (for reasons which could be spelt out here, but will not be!). In addition, the approach of the project coordinator, who adopted a "we have ways of making you do action research" attitude, deterred a number of teachers. (In fairness to the coordinator, it must be said that the idea of creating an action research dimension to the project was entirely his doing, and without his work it would never have been thought of much less tried.) Thirdly, professional development in general, and action research in particular became caught up in a union dispute with the administration over conditions of service and teacher appraisal. Fourthly, teachers felt that if their project were not successful, that this would be "used against us". In short, local politics got in the way of pedagogy.

Happily, one successful initiative evolved out of the failed action research project. This was a programme of peer observation, where teachers paired up, observed each other's classes, and then provided feedback. At the present time this initiative is growing in strength. It is under the total control of the teachers involved, there is no reporting back to the administration and the additional time involved for briefing and debriefing is worth it in terms of the rewards to

both observer and observed (the observer gets new ideas for his/her own teaching, while the observed gets useful feedback on his/her performance).

What Went Wrong?

Despite the fact that these projects all took place in widely differing contexts and educational environments, they all shared a number of characteristics. All initiatives ran into trouble, and in all cases, classroom change was less tangible than had been desired. In all instances, in one way or another, several things happened that prevented change. These are set out and discussed below.

1. Teachers were not given recognition or time off, etc. for doing the research. The time required to do even the most limited research project is immense. The teacher or teachers concerned need to conceptualize their research project, to discuss it, formulate an action plan, gather or create materials, seek advice, transcribe interactions, analyze data, and write the project up. Failure to provide adequate (or even any) time carries more than a resource implication. It also carries the message that the administration itself does not value the efforts that teachers themselves are making to change, in a critical and informed way, what is happening in their classrooms. Even in those projects where limited time off was provided by the administration, the lack of recognition given to the additional efforts which teachers put into their projects led to a feeling of demoralisation.
2. The agenda was hi-jacked by the administration. A second factor which worked against the success of the projects was the fact that the administration interfered in the projects in one way or another. Even in cases where I believe there was no intention to interfere, there was a suspicion on the part of teachers that the process was either controlled by the administration or likely to become so. This suspicion was fuelled by comments from inside and outside the administration which were often misinterpreted or taken out of context, but which nonetheless had the effect of creating a climate of suspicion which helped to either destroy the project or seriously weaken it.
3. The agenda was subverted from within by teachers who wanted to bolster their own position within the political context in which they worked. For reasons which are partly explicable, partly inexplicable, action research seems to be a lightning rod which attracts controversy and contention. In all three case studies which form the basis for this paper, there were individuals who wanted to use the existence of, or promise of, action research to advance their own political interests within the organisations

they worked for. Regretably, some teachers wanting to do research were criticised by peers as having either intellectual ideas above their station, or as wanting to curry favour with the administration. In one instance, the project director effectively subverted the project by being too directive and by alienating the teachers.

4. Teachers lacked the technical skills and knowledge to conceptualise and operationalise the research that they were interested in.

In all of the case studies reported here, the majority of teachers lacked the conceptual and technical skills needed for planning and implementing their chosen research. I believe that most university based academics (and here I include myself) seriously underestimate the skills needed to do this sort of work. Just as problematic was the perception of some teachers who, I believe, were eminently capable of doing action research, that it was "too difficult".

5. Doing research got in the way of teaching.

In all three cases described above, teachers reported that adding a data collection dimension to their work added considerably to the burden of teaching. The need, in most instances, to set up audio and video recording equipment and ensure that this was working, the problems associated with collecting data from students in pair and groupwork situations, and the worry that the data being collected would not feed into their projects caused a deterioration in their teaching.

6. Action research was partly associated with curricular innovations and there was an attempt to use action research so determine whether or not the innovation was leading to change in the classroom.

Each of the projects described in this paper was associated with an attempt at curriculum change. Motives varied from government decree and economic imperative to changing pedagogical fashions. In all cases, as soon as the participants sensed any form of intervention (read "interference") they retreated from involvement.

7. Teachers were fearful that involvement in action research would be used against them.

In at least one of the case studies described above, teachers were fearful that the action research projects would yield data that could be used against them. Several teachers reported that if they did not get positive answers to their investigative questions, then they would be labelled as "failures". Several teachers also felt that if their action research required audio and video recordings of their classes, these recordings could conceivably fall into the hands of the administration and be used for accountability

purposes. This, of course, is totally at odds with the spirit of action based inquiry.

What Went Right?

I would not like to conclude this piece by suggesting that action research was a total failure in the three cases concerned. While success was tempered by the factors outlined above, each project did exhibit some successes in prompting classroom change. In this section, I shall briefly outline a number of things that happened in the course of the projects that provide useful guidelines for future action. In the following points lie seeds for success. While the three cases which form the data-base for my paper would probably not be considered particularly successful by those on the inside, they do provide useful pointers for anyone wishing to initiate an action research agenda in their own context and situation.

1. Start small.

Those of us who have been involved in action research initiatives over many years now tend to underestimate the social and psychological pressures on those who take part in such projects, the practical complexities of collecting and making sense of data during the course of teaching, and the difficulties involved in conceptualizing research. In one instance, the teachers backed off from action research, but became enthusiastically involved in a programme of peer teaching, which shares several key characteristics with action research. Hopefully, in time, involvement in the peer teaching programme will lead beyond observation and analysis to intervention and experimentation.

The most successful projects were those where the teacher either started small, or managed to reduce the scope of their original idea to manageable proportions.

2. Negotiate with the administration and arrange trade offs, including time off.

While most administrations were notably reluctant to provide support in terms of paid time off, in cases where some modest recognition was made, this paid off in terms of teacher enthusiasm and involvement. It was as if participants simply wanted an acknowledgement that what they were doing was perceived to have value.

3. Get someone in from the outside to legitimate the exercise.

Anecdotal evidence suggested that the involvement of an external consultant, adviser or 'critical friend' assisted those involved in the process. The assistance might be technical help in selecting appropriate methods for data collection and analysis, or assistance in conceptualising the project.

However, the real value seems to be a psychological one - someone beyond the institution itself feels the project has value and is prepared to invest their own time and expertise. In addition, involvement of a credible external agent helps to legitimate the project to colleagues and administration.

4. Have someone on the inside to "own" the process.

There are no instances of action research that I am aware of that got beyond the initial discussion stage without an active leadership role being played by someone inside the organisation who was prepared to take ownership of the project. The nature of the ownership might vary, from negotiating with the administration to convening discussion groups to contacting external consultants. Regardless of their role, and irrespective of their status within the organisation, these individuals were crucial to the success (limited or otherwise) of individual projects.

In this presentation, I have chosen to look at three action research projects which were problematic, but within which were certain indicators of success. The characteristics set out in this section were critical to the limited success of the projects I have described. They reflect and reinforce a number of points which I set out some years ago in an introductory book on action research which are reproduced in Appendix 1.

5. Consider less threatening or less time-consuming alternatives.

I have already discussed the use of peer observation as a potentially valuable alternative to action research, which may nonetheless lead in the direction of action research.

6. Tell stories.

Another idea which, like peer observation, is gaining currency is the employment of narrative accounts whereby teachers tell their own stories in their own ways. (I hasten to add that the notion of research as 'storytelling' has been around from a long time. See, for example, Denny, 1978; Goodson and Walker 1982; Walker, 1982).

The notion of narratives as a justifiable form of documentation has been powerfully argued by Freeman (1994). He argues that teacher-research as storytelling has three tangible benefits. It allows for the articulation of personal theories, it highlights the social nature of teaching, and it results in the emergence of naturally occurring texts.

Freeman tackles head on the fact that "story telling" can, in fact, be a euphemism for downright lying. He seeks support for his views from the Nobel laureate Toni Morrison who, in her Nobel lecture states that, "Narrative has never been merely entertainment for me. It is, I believe, one of the principal ways in which we absorb knowledge." (Morrison, 1994: 7). Another telling piece of support comes from Freema Elbaz, who states that:

Initially, a 'story' seems to be a personal matter: There is concern for the individual narrative of a teacher and what the teacher herself, and what a colleague or researcher, as privileged eavesdroppers, might learn from it. In the course of engaging with stories, however, we are beginning to discover that the process is a social one: The story may be told for personal reasons but it has an impact on its audience which reverberates out in many directions at once. (Elbaz 1992: 5)

Action Research and Classroom Change

In many instances, despite the difficulties concerned, teachers did get to the point of formulating a research question, collecting baseline data, finding an innovation, and documenting the effects of implementing that innovation. In all of these instances, teachers reported the project had changed their classrooms, and, further, that it had changed them as teachers. When this was investigated more systematically, the following changes to professional practice were reported.

How has your teaching changed? Complete the following:

Since I have been doing action research, I find that when I teach I now ...

	More	About the same	Less
1. tend to be directive	1	14	10
2. try to use a greater variety of behaviours	16	6	0
3. praise students	15	10	0
4. criticise students	0	11	13
5. am aware of students' feelings	18	6	0
6. give directions	4	16	5
7. am conscious of my nonverbal communication	11	14	0
8. use the target language in class	19	6	0
9. am conscious of nonverbal cues of students	12	12	0
10. try to try to incorporate student ideas into my teaching	20	5	0
11. spend more class time talking myself	1	9	15
12. try to incorporate student ideas into my teaching	20	5	0
13. try to get my students working in groups	15	18	0
14. try to get divergent, open-ended student responses	14	10	0
15. distinguish between enthusiasm and lack of order	9	15	0
16. try to get students to participate	18	7	0

Conclusion

In this paper, I chose as my point of departure the deliberately provocative, but intentionally ambiguous statement "Why action research doesn't work". The paper is intended to show, not that action research is a lost cause, but that there are identifiable factors which can either vitiate the projects or render them ineffective. The data base of the paper is three action research initiatives that I have been involved in in several countries over a number of years. All three case studies illustrate the central point that it is impossible to divorce any one element, in this instance, action research and professional renewal, from the context within which it exists. I have also used the paper to illustrate other forms of professional renewal and inquiry such as peer observation and teacher narratives that are in harmony with, and can lead teachers into action research.

There is no doubt whatsoever, that action research can work, that it can be an effective instrument for bringing about professional renewal and curriculum change. However, at both conceptual and practical levels, it is more difficult to initiate and sustain than most current thinking would have us believe.

References

- Alderson, C. and M. Scott. 1992. Insiders, outsiders, and participatory evaluation. In C. Alderson and A. Beretta (eds.) *Evaluating Second Language Education*. Cambridge: Cambridge University Press.
- Allwright D. and K. Bailey, 1991. *Focus on the Classroom*. Cambridge: Cambridge University Press.
- Bailey, K. 1992. The processes of innovation in language teacher development: What, why and how teachers change. In J. Flowerdew, M. Brock, and S. Hsia (eds.) *Perspectives on Second Language Teacher Education*. Hong Kong: City Polytechnic of Hong Kong.
- Denny, T. 1978. Story-telling and educational understanding. Occasional Paper Series, No. 12, November 1978. College of Education, Western Michigan University.
- Elbaz, F. 1992. Hope, attentiveness, and caring for difference: The moral voice of teaching. *Teaching and Teacher Education*, 8, 5 - 6.

- Freeman, D. 1994. The storyteller, the anthropologist, the theoretician, and the social activist: Charting the dimensions of teacher-research. Paper presented at the TESOL France Convention, Paris, October 1994.
- Fullan, M. 1982. *The Meaning of Educational Change*. New York: Teachers College Press, Columbia University.
- Gebhard, J. and A. Ueda-Motonaga. 1992. The power of observation. In D. Nunan (ed.) 1992. *Collaborative Language Learning and Teaching*. Cambridge: Cambridge University Press.
- Goodson, I. and R. Walker. 1982. Telling tales. *Case Study Methods*, 4, 31- 56.
- Kennedy, C. 1987. Innovating for a change: Teacher development and innovation. *ELT Journal*, 41, 3, 163-170.
- Lewis, C. 1992. Action research with French immersion teachers: a pilot study. Unpublished monograph, University of British Columbia: Canada.
- Lomax, P. (ed.) 1989. *The Management of Change*. Clevedon Avon: Multilingual Matters.
- Markee, N. 1994a. Curricular innovation: Issues and problems. *Applied Language Learning*, 5, 2.
- Markee, N. 1994b. Toward an ethnomethodological respecification of second language acquisition studies. In A. Cohen, S. Gass, and E. Tarone (eds.) *Research Methodology in Second Language Acquisition*. Hillsdale N.J.: Lawrence Erlbaum.
- Mickan, P. 1991. LIPT: Languages Inservice Program for Teachers Stage 3 1990. Action Research Reports Volume 6, March 1991. Adelaide: Languages and Multicultural Centre.
- Morrison, T. 1994. *Nobel Lecture on Literature*. New York: Alfred Knopf.
- Nunan, D, 1989. *Understanding Language Classrooms: A Guide for Teacher Initiated Action*. London: Prentice Hall.
- Nunan, D. 1992a. *Research Methods in Language Learning*. Cambridge: Cambridge University Press.

- Nunan, D. (ed.) 1992b. *Collaborative Language Learning and Teaching*. Cambridge: Cambridge University Press.
- Richards, J. and D. Nunan (eds.) 1990. *Second Language Teacher Education*. Cambridge: Cambridge University Press.
- van Lier, L. 1988. *The Classroom and the Language Learner*. London: Longman.
- van Lier, L. 1992. Not the nine o'clock linguistics class: investigating contingency grammar. Unpublished monograph, Monterey Institute for International Training, Monterey: California.
- Walker, R. 1982. On the uses of fiction in educational research (and I don't mean Cyril Burt). *Case Study Methods*, 4, 57 - 81.
- Wallace, M. 1991. *Training Foreign Language Teachers: A Reflective Approach*. Cambridge: Cambridge University Press.

APPENDIX

Ideas for maximizing the possibility of achieving success in planning and implementing action research projects (Statements are from Kemmis and McTaggart 1988; comments in brackets are from Nunan 1989)

1. Get a research group together and participate yourself. (Comment: not only can sympathetic and like-minded colleagues be a useful source of ideas, they can provide a great deal of support when things get tough.)
2. Be content to start work with a small group. Allow easy access for others. Invite others to come when the topics that interest them will be discussed. (Comment: rather than trying to 'sell' the idea of collaborative research to sceptical colleagues, it is far better to work with a small group who are in sympathy with the idea of collaborative teacher-research. Once the group begins to work effectively together, the idea will sell itself.)
3. Organisationally, get things started by arranging an initial launch, identifying a nucleus of enthusiasts negotiating meeting times and the like. (Comment: the same remarks apply here as to point 2.)
4. Start small: offer simple suggestions to get people started. For example, who talks in your classroom and who controls the development of knowledge in your classroom group? Work on articulating the thematic concern which will hold your group together and establishing agreement in the group that the thematic concern is a shared basis for collaborative action. (Comment: as we have already seen, one of the errors that is often made by people undertaking classroom research for the first time is to set themselves research projects that are far too ambitious.)
5. Establish a time line: set a realistic trial period which allows people to collect data, reflect and report over two or three cycles of planning, acting, observing and reflecting. (Comment: the research process inevitably takes longer than initially envisaged, particularly once unanticipated problems begin to appear. Time needs to be allowed for these to be worked through. By drawing attention to the desirability of working through several research cycles, to ongoing rather than episodic nature of teacher-research is emphasised.)
6. Arrange for supportive work in progress discussions in the group. (Comment: once again, the value of collaborative support is highlighted. Regular work-in-progress meetings can also serve to reassure participants that progress is, in fact, being made.)

7. Be tolerant and supportive: expect people to learn from experience and help to create conditions under which everyone can and will learn from the common effort. (Comment: collaborative research is stressful because the participants are placing themselves in a situation where their real or imagined professional shortcomings come under scrutiny by others. For this reason, it is imperative that a supportive environment be created from the beginning.)

8. Be persistent about monitoring. Collecting compelling evidence is essential to ensure that people are learning from what their experience actually is. Be suspicious of claims made without evidence. (Comment: an important initial phase in the workshop outline presented in Chapter Six was to get participants to articulate their beliefs about language and learning, and then to produce evidence for these beliefs.)

9. Plan for the long haul on the bigger issues of changing classroom practices and school structures. Remember that educational change is usually a slow social process requiring that people struggle to be different. Change is a process, not an event. (Comment: implicit in the concept of teacher-research is the notion of change. We are doing more than simply observing our classrooms to see what is going on there. We are also intervening with the hope of improving on current practice.)

10. Work to involve (in the research process) those who are involved (in the action), so they share responsibility for the whole action-research process. (Comment: this point, once again, underlines the desirability of adopting a collaborative approach to research.)

11. Remember that how you think about things, the language and understandings that shape your action, may need changing just as much as the specifics of what you do. (Comment: one of the great benefits of involving others collaboratively in our research concerns is that it helps us identify and, hopefully, overcome the limitations of our own ways of thinking, acting and reacting.)

12. Register progress not only with the participant group but also with the whole staff and other interested people. Create a reputation for success by showing what is being done. (Comment: while it can be counterproductive to try to involve those members of the school community who are unenthusiastic about teacher-research, it is important to keep the community informed of what one is doing and the progress one is making. It is also important to avoid being seen as an exclusive, elitist group.)

13. If necessary arrange legitimising rituals, involving consultants or other 'outsiders' who can help to show that respected others are interested in what the group is achieving for education in the school. (Comment: this can be particularly useful if there is a feeling of scepticism among others in the school community who feel you are just playing at research.)

14. Make time to write throughout your project. Write at the beginning, during the project and at the various 'endings'. (Comment: the value of diary writing has been underlined here several times, but particularly in Chapters Four and Six.)

15. Be explicit about what you have achieved by reporting progress. For example, you can write up an account of your research project for others. Invite them to understand your educational theorising, to try the practices you have tried, and become part of the widening community of action researchers interested in the educational issues you have addressed. (Comment: the comments made in relation to points 12 and 13 are relevant here as well.)

16. Throughout, keep in mind the distinction between education and schooling. Action research is a concrete and practical process which helps those involved to build a critique of schooling, form the perspective of education, and to improve education in schools. (Comment: it is also important to keep in mind the value of teacher-research as an aid to one's own professional development.) (Points extracted and adapted from Kemmis and McTaggart 1988: 25-8.)

2

HELPING TEACHERS CHANGE THE LANGUAGE OF THE CLASSROOM: LESSONS FROM IN-SERVICE TEACHER EDUCATION

*Philip Hoare, Stella Kong
Hong Kong Institute Of Education*

Introduction

In September 1994 the Hong Kong Government began implementing major changes to its medium of instruction policy in secondary schools. This policy, if it is carried through, will have far-reaching implications for all levels of education in the territory. In February 1994 the then Institute of Language in Education, now a part of the Hong Kong Institute of Education (HKIEd), began its first full-time in-service course for teachers in secondary schools who use or intend to start using English as the medium of instruction (EMI). This course is the only one of its type in Hong Kong. We are now at the end of the second course. The course is open to teachers of all subjects taught through English, including English as a subject. It provides help to teachers who are accustomed to using mixed code in the classroom and now need to use English consistently. It also helps English teachers consider how they can provide support to teachers of other subjects in an EMI school. The course does not try to persuade teachers or schools to choose English medium education over Chinese medium. It is with this course in mind that we are approaching the topic of this paper. Teachers attending the course have also shared their views with us and these are reflected in the paper.

Our purpose, then, is to set out what we have learnt over the past year about the areas in which EMI teachers in Hong Kong need support. We will outline the most important qualities needed by an EMI teacher and will then describe how the HKIEd course tries to help teachers improve in those areas. We will discuss teachers' reactions to the course and the developments that are taking place as a consequence of this. Lastly, we will discuss other forms of support that are needed for EMI teachers in Hong Kong.

Qualities of English Medium Teachers

1. Language Proficiency.

It comes as no surprise that subject teachers we have worked with believe that an improvement in their language proficiency will lead to the most immediate and

significant improvement in their ability to teach through English. Johnson, Shek and Law (1993) mention language proficiency as one of three important qualities needed by such teachers. Swain (1991), in noting the importance of avoiding the mixing of languages in the classroom and the importance of linguistic feedback, implies a considerable level of language proficiency. Bernhardt and Schrier (1992) found "native or near-native fluency in the target language" to be a prerequisite for immersion teachers in the United States. Teaching through English to students with a restricted command of the language requires the ability to deliver content that is at the students' cognitive level, through English that is at their language level, at a time when there is a major mismatch between the two. As appropriate teaching materials may not be available then teachers must be able to produce or adapt this content in written as well as in spoken English. The level of proficiency regarded as necessary is a complex and probably controversial matter. Chan (1992), did attempt to specify a level in terms of the language functions teachers need to use in lessons in Hong Kong schools but her framework was never piloted. Johnson, Shek and Law (1993) emphasise the importance of fluency in the classroom, while acknowledging that accuracy is "a good target to aim at" (p31).

2. Teaching Skills and Strategies.

The second quality required by EMI teachers is a set of skills and strategies for teaching through a second language. Johnson, Shek and Law (1993) suggest that these are "specific to English medium teaching" (p28). On the other hand, the authors of a useful series of handbooks for immersion teachers produced in Montgomery County, Maryland (Montgomery County, 1989) suggest that these strategies "are similar in nature but differ in intensity and duration from strategies and techniques used by non-immersion teachers". An immersion teacher must accept responsibility for teaching subject content and developing language proficiency in the same lesson and perhaps at the same time. This requires the control of a range of planning and teaching strategies and skills that put far greater emphasis on language than subject teachers are accustomed to doing without, in the long term, detracting from the attention given to subject knowledge. Teachers must be able, for example, to explain subject content through English and anticipate language difficulties in content topics. They must provide both language support and opportunities for students' use of English, together with appropriate feedback.

3. Attitude and Background Knowledge.

The third quality, or set of qualities, required by teachers who wish to change to the use of English as the medium of instruction is a genuine commitment to the change. The language of instruction in schools is a controversial topic. As there are elsewhere, there are strongly held views on all sides in Hong Kong and those

views, quite rightly, represent political, social and economic priorities as well as educational ones. Teachers are under pressure from the school, from their students some of the time, from Hong Kong society and from educational tradition. Under these circumstances the attitude and commitment of the teachers towards the language policy they are working within is going to be a crucial determiner of their success.

This commitment is made up of several parts. Firstly, English medium teachers must understand the theoretical background for what they are doing in order to be satisfied that it can be effective. Secondly, they must be convinced of the benefits of the change. These benefits may be seen as educational, social, economic or political and individual teachers may prioritise them differently. If teachers question the appropriacy or effectiveness of using English then the temptation to give in to the pressure to emphasise the transmission of "key facts" over the use of language to develop understanding of content is likely to become too great. The mixing of English with Cantonese will then increase. The teachers' own doubts will then be confirmed as the students get inadequate exposure and practice, and can, indeed, only understand and express their understanding of the subject in Cantonese, if at all. Thirdly, the individual teacher must believe that she can implement the changes in her own classroom that will enable students to learn her subject effectively through a second language. This implies confidence in her linguistic ability and her ability to control the classroom skills and strategies described above. Lastly, commitment is only likely to be maintained in the long term if teachers receive the consistent support of colleagues, school management and the educational establishment.

The HKIED In-Service Course for EMI Teachers

Any course that proposes to help teachers to change their classroom behaviour in the way that we have set out to do must address these issues of language proficiency, skills and strategies for implementing English medium teaching, and attitude. Broadly speaking, then, these represent our course objectives. To date we have worked with 84 teachers. If the HKIED continues to support the course, then numbers are likely to rise substantially. Currently, our course is composed of 3 modules: the Medium of Instruction Module (MIM), the Language Improvement Module (LIM) and the Project Module (PM). The 3 modules, running simultaneously, each feeding into and supporting the other two, are designed as an integrated whole in working to achieve the aims of the course; they exemplify the relationship between language and learning. The units in the MIM discuss issues concerning the implementation of EMI and provide input on strategies for effective EMI teaching which teachers have to carry out in their projects. In the LIM, teachers work on various aspects of the English language

(e.g. pronunciation, vocabulary development) to improve their proficiency and strengthen their confidence in and capability of using EMI. The PM gives teachers chances to put into practice what they have learnt in the MIM and to use EMI in the classroom. Language improvement also results from the use of English as the formal and informal medium of the course. There is no formal assessment of the language proficiency of teachers on the course. This is impossible at present because, as we have discussed above, we lack a recognised threshold level of English proficiency for EMI teachers. We consider it potentially damaging to the teachers to have any such assessment on the course in the absence of any recognised qualifications for EMI and/or CMI teaching.

Evaluation of the Course and Teachers' Perceptions

The information presented in this section is taken from questionnaires, discussions and informal feedback from course participants, together with the views of academic staff working on the course.

1. Improvement of Language Proficiency

So far, almost all the teachers who have been on the course were able to sustain the use of English in the classroom, though with far from complete accuracy. This was the first challenge presented by their school-based project work, and most met the challenge. In view of the fact that this may have been the first time these teachers had really used English rather than mixed code to teach, it must be considered a success. The experience was also helpful in terms of the confidence they gained. When asked to state 2 things they found most valuable about the projects in the end-of-course evaluation questionnaire, about a quarter of the teachers on the first course mentioned either teaching in English or confidence in using English.

Despite this, many teachers did not believe that their English was good enough to use as a medium of teaching and they consistently expressed worries about being "bad models" to students. The aim of improving teachers' language proficiency was rated as that which was the least successfully achieved. This feeling is backed up by our own observations. Though the course may have helped to boost their confidence, it may not have been as successful in helping to improve their accuracy. A lot of them, for example, could not pronounce some of the content words of their subjects correctly and many had problems with basic language structures. The extent of the difficulties experienced by the teachers with, for example, question forms is worrying (for example, "What this mean?" instead of "What does this mean?").

Teachers are expected to use English at all times during the course for all purposes, both with tutors and among themselves. Though about half of the teachers on the first course considered that their spoken language had improved most, many did not speak English among themselves. They did not feel comfortable doing so. Yet English teachers on similar programmes do so with few reservations. This may imply something about their respective attitudes towards the English language.

The teachers' knowledge about language and the part it plays in the learning of their subjects is also disappointing. When they were required to identify language objectives for teaching, they were only able to identify some very general language structures (for example, the passive voice, past tenses, prepositions) and/or vocabulary. These were frequently not significant for the subject or topic they were teaching. They had no concept of language functions or discourse features at all. There remains, therefore, a real question as to the extent to which they can integrate content and language in teaching their own subjects.

2. Skills and Strategies for EMI Teachers

Teachers did try out, to various extents, the "new" skills and strategies introduced to them on the course in the school-based part of their project work. They took more care to ensure that their own explanations of subject content given in English were at an appropriate language level. They provided more opportunities for their students to use English. Unfortunately there was a general lack of the language support that students needed if they were to produce the language required.

A common area of concern among the teachers was the feasibility of using these skills and strategies when they returned to school because of the time involved in preparation and the lack of support from other colleagues and the school. They did not see themselves as having the "power" to decide on what and how much to teach and, therefore, felt unable to make the necessary reductions in syllabus content in the early stages of using English. In their school-based work, most of them included the same amount of subject content as they usually do in their normal lessons in school. They therefore failed to finish their teaching plans since they were also trying to include more attention to language. The teachers were particularly worried about exams, particularly public exams. If exams do not demand the effective use of language by students, they were not prepared to do so either, however important they may recognise it to be.

3. Language Across the Curriculum Strategies

Teachers on the course were introduced to a range of basic language across the curriculum strategies. These included the use of the same set of classroom

language, the same expectation of the level of language production from students, regulating the amount of vocabulary across the subjects for students, and identifying links of content and/or language across subjects. They were also expected to put these into practice in their school-based work. To enable them to do this, a group of 4-6 teachers went to the same school to teach the same class for 3 days. While they each taught their own subject, they were encouraged to integrate the use of English across subjects.

It was, of course, not easy to assess the success of the teachers' efforts in trying out these strategies within just 3 days. However, teachers on the course did discuss at length the feasibility of implementing them in school. Their concerns were the same as those described in the last section. They expected an increase of workload; they believed that it was impossible to work alone and that there was a need for the whole school to collaborate. But again, they did not believe that they had the power or the ability to persuade their colleagues to cooperate. Some of the teachers did not believe that cross-curricular co-operation on language in learning could be of any value.

4. Relationship between Language and Learning

Despite the fact that more than half of the teachers on the first course gave a very high rating for the extent to which this aim was achieved, teachers generally had serious problems in trying to integrate language into subject content teaching. Some of them insisted on viewing the roles of subject teachers and English language teachers as distinct. They could not accept any responsibility for the development of their students' English in the classroom. Some teachers refused to identify language objectives, beyond vocabulary, when planning their projects.

In their evaluation of the course, quite a number of teachers commented that the course should concentrate only on their language improvement, ignoring classroom strategies. They often complained, quite rightly, that there were no subject specialists among tutors on the course and that this compromised the effectiveness of the course. It is certainly true that more subject-specific expertise among course tutors would be useful though at present it is not available. Nonetheless, refusing to admit the importance of language in subject content because it is put forward by language experts indicates a failure to see the relationship between language and learning which we were seeking.

Directions for Course Development

While recognising the achievements of the course, we are also aware of areas where improvement is needed. Some of these are of an organisational nature and, while they are vital, they are not relevant to this discussion. The most important

areas that we must strengthen are those that relate most immediately to the special role of an immersion teacher in teaching subject content through a second language. Some of these areas are at a conceptual level, others at a very practical level.

First, there is the need to develop more subject-specific materials for both the MIM and the LIM units in order to strengthen the teachers' understanding and acceptance of the importance of language and content relationships. The range of materials on grammar and discourse and the pronunciation of subject specific terms will be extended. In order to increase the emphasis placed on individual subjects, the planning, production and school-based work of one project will in future not require cross-curricular collaboration and will be based in the teachers' own schools. We still do not propose to run separate courses for teachers of individual subjects. We plan, also, to provide much more background knowledge of language, which we originally believed might be resented. This will develop a better understanding of the relationship between language and content.

The second area of development is collaboration between the EMI participants and English teachers attending similar in-service programmes in the Institute. We have introduced a sharing session for these teachers so that they could identify areas of mutual support and collaboration. We are now looking into the possibility of integrating some of the units of the two courses to bring about a closer understanding. We hope to raise levels of understanding and co-operation both while the teachers are on course and when they return to their schools.

A last area where we see the need for development of the course is the introduction of formal assessment of language proficiency for teaching purposes of teachers on finishing the course. This is a long-term target and, as has been mentioned, this can only be meaningfully done if there is a similar assessment on pre-service courses.

Support Beyond in-Service Provision

The well-known limitations of in-service programmes and the number of teachers requiring help suggest that the HKIED course is inadequate as the only or even the chief support for teachers in changing their medium of instruction. There is an additional problem. A good many of our teachers, as described by Hoare and Lee in another paper in this conference, are cynical about the commitment of the government to its own medium of instruction policy. They are convinced that, however much better equipped they may be to implement it, they will not be permitted to work in the ways they know to be necessary.

If teachers are to overcome this feeling of helplessness there are five main areas in which they need support. The first of these is teacher education. The further development of in-service provision has already been discussed. A similar commitment at the pre-service stage is equally necessary. New teachers of all subjects must be aware of the differences between teaching through the mother tongue and through a second language. This distinction needs to be maintained through the teaching medium of their course, the nature and amount of English language provision, through the work they do in their specialist subject areas and in the assessment of the course.

Better understanding and support from within schools is essential. Many principals still only pay lip-service to the medium of instruction choice they have made. Many do not understand or cannot accept the changes that are needed. They are certainly not encouraging teachers to implement changes. Implementation frequently stops at exhortations to use more English. There are schools that are doing more but too few of them. Above all, leadership is required. Properly co-ordinated in-school staff development over a substantial period is essential. A willingness to reconsider textbooks, school examinations and subject autonomy is also needed.

We cannot, however, fail to appreciate the pressure school managements feel they are facing. They, in their turn, need practical support from the government and the education system, in a form that goes beyond the pronouncement of a policy and some minimal guidance. This should include a much more concerted attempt to inform parents of the issues and help in adapting syllabus content to take account of their chosen medium of instruction. At present, schools seem to have no confidence that the new medium of instruction policy will be seriously supported.

Fourthly, there must be further changes to the examination system, which will demand the use of language to demonstrate an understanding of the subject and not simply to list facts. This process has begun but will have to go much further if it is to influence teaching.

Lastly, much more local research is needed. The research on which the current approaches are based comes from different education systems, with different languages and, crucially, different educational and social values. While theory transfers across contexts, the nature and implementation of a programme is dependant on the context. We have very little hard information to give teachers about what will work in Hong Kong classrooms and the extent of the changes they must make.

Conclusion

Bringing about classroom change is in large part the responsibility of the classroom teacher. We have tried to show how teachers trying to implement change in their medium of instruction within Hong Kong's language in education policy are being helped through in-service education. We have also indicated some of the problems this has given rise to and how these are being addressed on this course. Lastly, we have briefly described the other forms of support that are essential to teachers if they are to implement the change successfully. While in-service education can be an important contribution it is insufficient on its own. Teachers can be agents of change but only within a professional environment that supports and values their contributions.

References

- Bernhardt, E. and Schrier, L. 1992. The Development of Immersion Teachers. In: Bernhardt, E.B. *Life in Language Immersion Classrooms*. Clevedon: Multilingual Matters.
- Chan, W.M. 1992. *Minimum English language Requirements and English Courses for College of Education Students: Research Report*. Hong Kong: HongkongBank Language Development Fund and Institute of Language in Education.
- Johnson, R.K., Shek, C.K.W. and Law, E.H.F. 1993. *Using English as the Medium of Instruction*. Hong Kong: Longman.
- Board of Education of Montgomery County. 1989. *What it Means to be an Immersion Teacher*. Maryland: Montgomery County Board of Education.
- 1989. *Negotiation of Meaning*.
 - 1989. *Planning for Instruction in the Immersion Classroom*.
 - 1989. *Foreign Language Immersion: an Introduction*.
- Swain, M. 1991. The Immersion Experience in Canada: Is it Relevant to Hong Kong? In: Bickley, V. *Where from Here? Issues relating to the Planning, Managing and Implementation of Language Teaching and Training Programmes in the 90's*. Hong Kong: Institute of Language in Education.

3

ASSESSMENT AND CHANGE IN THE CLASSROOM

Vivien Berry, Peter Falvey, David Nunan, Martin Burnett and John Hunt¹

Current Assessment Issues and Practices in Hong Kong: A Preview

Vivien Berry

English Centre, University of Hong Kong

The aim of this colloquium is to focus on a range of issues which are of major interest to all participants within the educational system in Hong Kong. The following papers by Peter Falvey, David Nunan and Martin Burnett and John Hunt, discuss assessment procedures either currently in place or in the planning stages for all ages ranging from children in their first year of primary school through to students graduating from tertiary institutions. The implications of some of these assessment procedures for selection into (and rejection from) progressively higher levels of education and the potential impact they will have on the lives of literally hundreds of thousands of young people over the next few years are of the utmost importance to each and every stakeholder within the educational system.

One of the major issues within the field of assessment in the 1990s has been a concern with the systemic validity of tests - the so-called "washback effect" or the effect a test has on classroom practice. In their extremely thorough and useful record of current English language testing practices in Hong Kong, Boyle and Falvey (1994:xi) even go so far as to suggest that the washback effect has "...joined validity, reliability and practicality as one of the Big Four considerations in evaluating the worth of a test." (see Bachman, 1990, and Bachman and Palmer, forthcoming, for the most comprehensive discussions of how to determine test usefulness currently available). Whether or not a test, or series of tests, should drive a curriculum is a matter of ethical concern which can not possibly be addressed here; interested readers are referred to Shohamy, 1992 for a powerful discussion of the uses and abuses of tests within second language educational contexts. Nevertheless, there does appear to exist within the

¹ The papers in this chapter were presented as a colloquium in which Professors Lyle Bachman and Bernard Spolsky also participated as discussants

collective minds of testing bodies throughout the world, something akin to what King (1994:11) describes as "... a testing cliché that underpins the whole of the Hong Kong examination scene: *"If we don't test it, they won't teach it!"*".

This *"us and them"* view of the purpose of assessment is perhaps best (and most generously) viewed as a skills-based, as opposed to a knowledge-based, approach and is probably most often found within English language testing circles. In other words, if a test has a listening skills component, a teacher will make sure the students get sufficient practice at listening. However, Andrews (this volume) and Lam (ibid) examine some of the effects of the introduction of the oral component into the 'A/S' level Use of English Examination, concluding that the cause and effect relationship is much more complex than originally assumed. Outside language testing circles, it is very hard to imagine an examination authority taking a unilateral decision to introduce items into, for example, an 'A' level biology test that had not been previously specified by subject specialists in a biology syllabus. Such an autocratic and authoritarian approach to assessment, however well-meaning, would, quite rightly, be the subject of instant howls of protest. This is not, of course, to suggest that "knowledge" and "skills"; are mutually exclusive; taken together with "performance", they form the collective basis of the constructivist approach to criterion-referenced educational assessment which is being introduced (often with considerable initial reluctance on the part of teachers) in school systems throughout the developed world.

These new forms of assessment, focusing on recording achievement rather than lack of it, saying what an individual student can do, rather than taking a norm-referenced view of how well an individual performs in relation to the entire peer group, will be introduced into the Hong Kong school system in 1995, albeit quite deliberately without any mention of the word "assessment". Peter Falvey's paper discusses the rationale underlying the recent exercise in curriculum reform in Hong Kong and also describes in detail how the Targets and Target-Related Assessment initiative evolved into the much more user-friendly-named Target Oriented Curriculum.

In addition to encouraging a criterion referenced approach to assessment, a result of the current search for alternatives to large scale proficiency tests and the involving of learners "...in discovering what they know and what they feel as well as what they can do." (Cram, 1995:271), has led language assessors to revisit the issue of learners' self-assessment. Much of the previous research within the context of self-assessment has been concerned with matters of reliability and validity or how well learners' assessments of

their ability match teachers'; findings in this respect have often been contradictory (cf. Rolfe, 1990 and Wilkes, 1995). Rather than being used merely as an alternative to formal assessment procedures, however, one of the major advantages of self-assessment is that if it is introduced into the classroom setting by a sensitive teacher, it can be a major tool in the development of learners' awareness of their learning processes and achievement. In other words, the great strength of self assessment may lie not in its ability to accurately assess learners' proficiency, however reliably, but in its ability to promote learner autonomy by encouraging self-reflection about the process of learning itself rather than the end-product. The preliminary results of the study presented in David Nunan's paper show that this is indeed an extremely promising area for further research.

The final paper in the colloquium is premised on the belief that accountability in education goes beyond the immediate stakeholders such as students, parents, funding bodies, etc. and embraces those who will have a future interest in educational outcomes, namely potential employers. As the change in sovereignty of Hong Kong draws inexorably nearer, anxiety over Hong Kong graduates' language ability has grown to embrace not only English but Cantonese and Putonghua as well. Without becoming embroiled in the debate over possible falling English standards, and the reader is referred to King (1995) for an absolutely categorical refutation that standards have fallen over the last twenty years, it is extremely easy to see why the impending transfer of control from the United Kingdom to the People's Republic of China is causing great concern amongst employers over their potential employees ability to communicate effectively in written and spoken Chinese. Quite simply, potential employers are looking to tertiary institutions to provide them with an easy-to-interpret, reliable measure of potential employees abilities in all specified languages. However, providing such a measure is not a straightforward task; there are an almost infinite number of variables to be taken into account when attempting to develop an assessment instrument which will be used to make decisions that will literally make or break someone's career. Martin Burnett and John Hunt's paper describes the origins and current status of an inter-institutional project, funded by the University Grants Commission, which is charged with developing instruments to assess graduating students' ability in all three languages. Obviously they can only refer in passing to the various developments in the Chinese sectors; their concern is with the English language assessment exercise. Refreshingly, and perhaps surprisingly in light of the examination-oriented nature of Hong Kong education to date, they argue against what they term a "monolithic exam-based model" in favour of a more interactive form of assessment which has

both active input and approval from all the involved participants including students, educationalists and the business community.

The Target Oriented Curriculum - Issues in Assessment

Peter Falvey

Dept. of Curriculum studies, University of Hong Kong

Introduction

In this paper, the Target Oriented Curriculum initiative and its implications for assessment will be discussed within the context of:

- * current learning theories
- * the differing roles of assessment in national school systems
- * developments in assessment in the Hong Kong education system
- * the implementation of TOC and its assessment practices in Hong Kong

Current learning theories

In the past, structuralist approaches to theories of learning predominated. Central to these theories was the assumption that knowledge could be broken down into atomistic "bits". These "bits" could then be learned, assimilated, regurgitated (to indicate that learning had taken place) and eventually applied.

Over the past few decades, as Educational Psychologists and other researchers have continued their investigations into how students learn, it has become generally accepted that learning does not take place in a linear, atomistic fashion but that knowledge is "constructed" by learners. This construction of knowledge occurs when students begin to fit new knowledge into their existing knowledge systems and create (or construct)

new meanings for themselves, This approach to describing how learning takes place has become known as a constructivist theory of learning.

When we examine the differences between these two theories of how knowledge and skills are learned, it becomes clear that perceptions of the roles both of the teacher and the student will differ considerably. In the structuralist view of learning, the teacher "provides" both content knowledge and the knowledge required for the acquisition of skills. The role of the student, in this model of learning, is to receive, learn and apply the knowledge and skills which have been provided for him. In the constructivist view, the teacher facilitates (or assists) the learning which is constructed by the learner him/herself. In the latter view, the learner takes responsibility for the acquisition and retention of knowledge and skills. He/she becomes an active rather than a passive learner.

Proponents of the constructivist view began to advocate new curriculum and methods of facilitating learning which focused more on the role of the learner at the centre of the learning experience than as a passive recipient of knowledge which is provided by the teacher.

Differing Views of the Purposes of Schooling

As Biggs (1995) has pointed out, as educational systems have changed so the role that assessment plays in schooling has altered. Biggs maintains that in the early stages of the development of schooling within an educational system, the system is often poorly researched. Many countries in South America, South East Asia and Southern Africa were, until relatively recently, in the early stages of economic, social and educational development.

In these situations, assessment is used primarily for the purpose of *selection*. Full, compulsory, free education is not feasible at this stage of development because of its cost. It is thus necessary to use assessment to select the top students who will then be allowed to take up the few places that exist in the system for secondary and, possibly, tertiary education. Parents, teachers and students alike believe strongly that in such a situation, a focus on examinations, preparation for examinations, a great deal of practice of examination materials, and eventual success in examinations (both school and public examinations) leads to a successful career and eventual prosperity. As a result of the pressures that examinations bring to teachers, students and parents alike, assessment is perceived by all these stake-holders as threatening. It can, and often does

become anti-educational and counter-productive to educational innovation and exploration.

However, once countries have achieved the economic and social development which allows for full, free, compulsory education up to the end of secondary education, it is possible to view the role of assessment and assessment practices in a different light. Instead of being used merely for *selection*, assessment can be used for the purpose of *education*.

This means that the focus of assessment need no longer be on selecting a few from a curriculum which is designed for a small section of the community, those who join the elite because they have been selected. Instead, assessment can be used for the primary purpose of enabling individual students to maximise their own learning potential. This can be done by providing continuous feedback:

on the progress of students,
on the effectiveness of learning programmes
for diagnostic purposes
for accounting for and recording what students have achieved
rather than what they have failed to achieve.

In short, the focus can be placed on the individual student and his/her own achievements. Instead of using assessment in order to deselect the majority of students, it can be used to report how well students have learned knowledge, skills and attitudes and how well they are able to interact with their peers and the world. It is intended, in such an educational context, that the recording and reporting of achievement stresses what has been achieved rather than what has been failed or unaccomplished.

These changes, however, bring their own sets of problems with them. Problems occur when the experts, educationalists, curriculum planners and assessment specialists, having become convinced that the old system is out-dated, begin to create a new system of curriculum design and assessment practices. The major problem occurs with the majority of people's perceptions of these changes when they do not initially share the convictions of the curriculum change-agents. The perceptions of most of the parents and many of the teachers, who are used to the old system, remain, at first, unchanged. They are used to dealing with the known, with an assessment system which brings its own pressures but which they feel they know how to deal with. Changes are resisted because they are unknown, uncertain and, therefore, potentially threatening to the system in

which the parents and teachers have been brought up. Hong Kong is currently undergoing that period of change in curriculum development and changing assessment practices.

Outcomes of constructivist theories of learning and a different approach to assessment

The movement towards constructivist learning theories coupled with a different approach to assessment, (based on criterion-referenced approaches to assessment rather than norm-referenced approaches to assessment) led to new approaches to curriculum design in different parts of the world (e.g. Scotland U.K. and Victoria State, Australia). Other countries later joined this approach (e.g. England and Wales). In these approaches the concept of performance tasks (or targets) as the basis for learning was proposed. Students would be informed clearly what it was that they were expected to achieve. They, and their parents would be able to read about and be shown examples of the performances, skills and knowledge which they were expected to cope with. In addition, these skills, and aspects of knowledge would be presented in "Key Stages" based on what the majority of "average" students would be expected to be able to achieve at that level. Study of what the average student could achieve would be represented in "Bands of Performance" within the Key Stages.

It was asserted that in this system of assessment, students would often be in the position of being able to declare to their teacher that they were able to achieve a target even *before* the teacher realised that it had been achieved. The reason for this would be because statements of Targets would be available to all. They would be accessible by parents, teachers and students. These skills and knowledge targets would no longer belong to the secret knowledge of the teacher and the principal but freely available to students, parents and the public. As a result it was hoped that students would be able to see clearly what had to be achieved and thus be able to report success rather than failure. Records would be presented in such a way that what a student could do would be reported rather than what he/she could *not* do. (Clark et al 1994: 14-15).

Experience in those countries which attempted this approach varied. As with any curriculum initiative, relative success or failure was often seen as a matter of whether the initiative was perceived as "bottom-up" (emerging from the teachers) or "top-down" (dictated by government).

Developments in Hong Kong

Within the Hong Kong educational system, educational planners began, after the publication of Education Commission Report number 4 (1989) to plan and implement a new curriculum and the assessment practices that it would require. The curriculum would be target-based and assessment would consist of assessing achievement of the targets to be set. It was named, by Education Department, Targets and Target-related Assessment (TTRA). The initiative met with an extremely hostile reception, mainly because the educational community and the general public perceived the initiative as top-down, hierarchical, and based on lack of consultation and briefing. Sengupta (1989) chronicled the early days of this initiative and predicted that the initiative would fail, not because of intrinsic lack of merit but because of the presence of most of the factors that lead to failure in curriculum initiatives (see Morris 1989, 1995). In addition, much of the documentation which accompanied explanations of the initiative was obscure and theoretical.

As a result of pressure from the educational community in which fear of changes predominated, particularly in assessment practices, government attempted a fresh start by changing the name of the initiative to "Target Oriented Curriculum" (TOC). In fact, little changed in the conception of the initiative from the days of TTRA. However, much greater attempts were made to educate the change agents (teacher educators of teachers attending orientation seminars to explain the changes, principals of primary schools and key teachers in the schools which would first implement the changes).

The Hong Kong Model

The model which emerged in Hong Kong has the following features:

1. Key Stages

Primary and secondary schooling would be divided into four Key Stages as follows:

Key Stage 1 = Primary 1-3

Key Stage 2 = Primary 4-6

Key Stage 3 = Secondary 1-3

Key Stage 4 = Secondary 4-5

2. Subjects to be focused on initially

TOC would be introduced in 1995 in Mathematics, Chinese and English.

3. Assumptions about subjects

Knowledge of the three subjects (Mathematics, Chinese and English), acquired in the 4 Key Stages would be categorised meaningfully into assessment dimensions as specified in the Programmes of Studies for each subject. The assessment dimensions were to be: knowledge, thinking, skills and attitudes. These would be assessed in eight Bands of Performance (BoPs) over the four stages.

4. Assumptions about learners

Progress along the learning path will be one of continuous development and growth. Learners will progress continuously from Band 1 to Band 2 and then on to Band 3 etc. (i.e. a learner at Key Stage 2 with Band 4 performance in English will have progressed through Bands 1, 2 and 3 of that subject.

Problems to be overcome in the area of assessment

1. Assessment practices

The move away from norm-referenced assessment to criteria-referenced assessment has to be explained to and understood by parents, teachers and principals. Parents are used to their children being ranked in all subjects in school. They see their children in competition against others for much-valued places in primary, secondary and tertiary institutions. Even while accepting the principles embodied by TOC they will need to be reassured that their children will not be disadvantaged in being selected for the primary or secondary school of choice.

2. Criteria for assessment

The following is a list of possible assumptions about criteria for making judgments about where the assessment points in the BoPs would be placed:

- criteria as the required knowledge and/or skills at specified exit points in the system
- criteria as the experts' opinions
- criteria as the norms of the target population
- criteria as the experts' opinions modified by the target group's performance
- criteria as the target group's performance scrutinized by experts

3 Specifications for task creation/construction

It will be necessary to create detailed specifications for assessment task writers. Such specifications are required because a task written by one writer to an unspecified brief may well be much more difficult or easier than a task written by another writer following the same inadequate brief. Parallel tasks must have clear specifications for each task writer so that the tasks become valid instruments for assessment.

4. Task banking - matching assessment to learning

It is important that a very large number of tasks be banked and made readily available to teachers. Fears about security should not exist because a valid learning task should resemble an assessment task. Assessment tasks should not be any different from learning tasks because they are assessing the same form of learning at the experiential, knowledge or interpersonal levels. If assessment tasks are matched to learning the validity of criterion-referenced assessment becomes clear. What the student learns is what is assessed and the learning to take place is specified clearly for all.

5. Recording and reporting - clarity and implications for teachers' workloads

One of the criteria for successful implementation of TOC and its assessment practices will be the clear, unambiguous recording and reporting of student progress through the BoPs. An imperative for the assessment experts is to devise recording and reporting formats which are clear, and accessible to and understood by all the stake-holders (parents, teachers, principals, students and officials). Another criteria for the successful implementation of TOC assessment is that it will have to be demonstrated to teachers that their workloads will not increase dramatically under the new assessment procedures.

6. Timing

Spolsky (1994) maintains that a curriculum initiative would normally take seven years to gestate, be piloted, evaluated and eventually implemented. Hong Kong has a proven track record in commerce and industry for adapting to and implementing change quickly and efficiently. Education reform, however, should not be equated to changes in industry and commerce. Adequate time is required for the education and orientation of parents and teachers. One of the problems with the proposed assessment changes is that implementation is scheduled to take place quickly, possibly before the public have had time to assimilate the changes (Sengupta op. cit.).

7. Evaluation and feedback

The Hong Kong government, especially in education, is not normally known for commissioning independent evaluation studies. Most evaluation takes place in-house. It will be important for the success of this curriculum initiative that government allows long-term, continuous independent evaluations to take place and that attention is paid to the feedback that such independent evaluations will produce so that its curriculum and assessment practices can be monitored and amended where necessary.

Conclusion

For any curriculum initiative to be successful it must be embraced by the teaching profession. Changes which start at the top will only be successful if teachers can see clear advantages for their students and themselves in such changes. They must be credible (at the level of theory and operation), understandable (conceptually and descriptively) and practical (at the level of recording and reporting). The next few years will show whether these criteria have been met when the effectiveness of the changes are evaluated.

Self-Assessment as a Tool for Learning

David Nunan

English Centre, University of Hong Kong

Background

The value of involving learners in their own learning processes has been well established in the literature. It is beyond the scope of this paper to review the literature in detail. The interested reader is referred to Oskarsson (1978), who argues that the education of the learner to make reliable and valid autonomous judgments on the effectiveness of his communication is a necessary part of the learning process; Nunan (1988) who provides a rationale for self-assessment as an essential part of any form of learner-centred instruction, and Underhill (1987) who argues that, in terms of self-assessment, it is the learner who is in the best position to

say how good he is at speaking on the basis that "he (sic) has been present at every effort he has ever made to communicate in the foreign language, while oral test assessments are based on a sample ten minutes' speech" (Underhill 1987:22)

Research questions

The action research study reported here took as its point of departure the following questions:

- Does guided reflection and self-reporting lead to a greater sensitivity to the language learning process on the part of the students?
- What effect does guided reflection and self-reporting have on the development of learning skills?
- To what extent does guided reflection and self-reporting lead learners to formulate more realistic learning goals?

Data collection instrument

The data collection procedure in this study was quite straightforward. Every two weeks students were given a blank pro forma containing the following probes. Students completed the form in their own time and submitted it at the next class.

PLEASE COMPLETE THIS FORM BETWEEN NOW AND NEXT WEEK
NAME: _____ UNIVERSITY NUMBER: _____
DATE: _____
This week I studied:
This week I learned:
This week I used my English in these places:
This week I spoke English with these people:
This week I made these mistakes:
My difficulties are:
I would like to know:
I would like help with:
My learning and practising plans for next week are:

Preliminary results

Preliminary results are promising. The following table contrasts comments made by students at the beginning of the course with those made at the end of the semester. It is clear from the data that by the end of the semester students were able to make much more detailed and accurate statements of what they wanted to learn and how they wanted to learn it. They were also better at making connections between the input they received in their English classes and their regular subjects.

Probe	At the beginning of the course	At the end of the course
This week I studied:		
	The nature of verbs	I read a journal called Geographic which is published in New Zealand. I have spent an hour to discussion with my psychology classmates.
This week I learned:		
	Some more information about English in English linguistics lesson.	The principles of morphology. How to use the self-access centre for learning English.
This week I used my English in these places:		
	Tutorials My German lessons.	In the library, Knowles building, KK Leung building. At home. Along the street near my home.
This week I spoke English with these people:		
	history lecturer, EAS classmates and tutor, linguistics tutor	A foreigner - he asked me where is Lok Fu MTR station. The waiter in Mario restaurant.
This week I made these mistakes:		
	Using incorrect words.	I spent too much time watching TV while answering questions. I created a word "gesturally".
My difficulties are:		
	Lack of time.	Understanding the theme of a topic or article. Writing fluent English essays.
I would like to know:		
	How to improve my English.	The method that can improve both my listening and speaking skills.

I would like help with:

Dictionaries.

Ensuring I would spend some time on reading but not on other leisure activities.

Communicating with foreigners.

Watching foreign films. Human resources that can improve my language ability.

My learning and practising plans for next weeks are:

To talk more.

To speak up in class and to use English to ask about anything I don't understand in any of my subjects. To try to understand by explaining to my schoolmates some topics of the essay before writing it.

Conclusion

This study was carried out in order to investigate what happens when a group of learners are asked to reflect on aspects of the learning process over the course of a semester. While this is a preliminary study, certain conclusions emerge. In the first place, opportunities to reflect on the learning process did seem to lead to greater sensitivity to the learning process over time. In addition, learners were able to see the relevance of their English course and their regular subjects. In other words, they were able to make greater connections between English and content courses. Finally, the data show that learners developed skills for articulating what they want to learn and how they want to learn.

The Graduating Students Language Proficiency Assessment Project (GSLPA)

Martin Burnett and John Hunt

Department of English, Hong Kong Polytechnic University

Background

In April this year (1994) the University and Polytechnic Grants Committee of Hong Kong (the UPGC) accepted a proposal from the Department of English of the Hong Kong Polytechnic, now the Hong Kong Polytechnic University, to lead an investigation into the feasibility of assessing the communicative proficiency of all first degree students attending the seven institutions that it funds. This assessment of proficiency in English, Cantonese and Putonghua (Mandarin) is to be made towards the very end of a student's course and is to provide prospective employers with an up-to-date, clear indication of what a student can do with the languages as that relates to communication in the workplace. The UPGC handsomely funded the feasibility study to the tune of HK\$ 7m. (or about US\$1m.)

The funding was made available on the understanding that the project should be inter-institutional across all the seven, now designated, universities; should cover all three major languages of business communication in the territory; and should focus on providing employers with "can do" statements about students' abilities to write and speak.

UPGC's purpose in sponsoring the feasibility study, (apart from the ostensible one of providing employers with useful, valuable information on individual prospective employees), is to use such an assessment (should it be put in place) as one more tool (among the many it has developed over the last few years) with which:

1. to attack the commonly perceived problem of the declining standard of English in the colony's workforce
2. to promote skills in Putonghua² which is seen as of rising importance in the colony's development
3. to focus attention on the need to raise generally that standard of native tongue skills in line with the changing requirements of the Hong Kong business community.

² According to the 1991 census Putonghua is spoken by only 1.1% of the population of Hong Kong as a first language and as a second by only 17%

One thing to be borne in mind right from the beginning of the study has been that the end-user of the assessments will be the business community and not language-orientated academics. The assessments should, therefore, not be of:

1. attainment of syllabus objectives,
2. potential for improvement, nor
3. knowledge of the languages.

The assessment should tell the employer what an individual, prospective employee can do now with the languages, in terms of Oral Interaction and Writing, that is relevant to the workplace.

Should such an assessment become, in one way or another, obligatory for all first degree students, and should the quite proper demand from the student body to be prepared for it be met, these two factors taken together would radically alter the current balance of at least the final year of degree studies for almost all those undertaking first degree courses in Hong Kong. The ability to communicate effectively in more than one language would become, quite rightly in our view, a major focus for all first degree academic programmes. The changes in the university classroom that will be required to align to this focus will be fundamental.

This two year project has been set-up initially to examine the feasibility of establishing the "machinery" to allow an assessment of the language skills of final year students to take place in late Spring of 1997. The number of graduating students is likely to be in excess of 15,000.

In this short paper we can only describe briefly what has been done to develop a framework for the testing machine and explain some of the issues that this project has brought to the surface and to offer suggested ways of addressing them. This presentation concentrates on the work done to date on the English language strand of GSLPA. Work on the Chinese languages strand, which is now well under way, is being conducted under the direction of Professor Cheung Yat-Shing, Head of the Polytechnic University's Department of Chinese, Translation and Interpretation and will be reported on separately.

Brainstorming and identification of major issues

So what has been done so far? After some ground clearing by way of hard-thinking and the sketching of some large and complex diagrams of implications inherent in the project, we visited all the departments, institutes and centres that provide ELT for degree students in Hong Kong for briefing/fact-finding sessions.

Last summer we were also lucky to be able to visit various language testing specialists in the United Kingdom to tap their thoughts.

Following on from this, we did two things: one was to write up a report on ELT provision in the seven institutions; and the second was to set up a working party made up of members from those institutions.

The Working Party came up with this list of issues that would need to be addressed:

1. The compulsory nature of the assessment
2. The nature of the communication to be tested
3. The form and substance of the collaboration between the UPGC institutions
4. The effect of an in-place test of language proficiency at the end of undergraduate programmes in 1997
5. The UPGC commitment to the development of the assessment and its possible consequences for language education in Hong Kong.

These are of course not entirely separate issues. They overlap and intertwine and feed off each other.

Problems and constraints

As for the compulsory nature of the assessment as contained in the original proposal, the feeling of the Working Party was that it would be difficult to make the assessment compulsory for all graduating students in Hong Kong at this stage, for a variety of reasons. There could probably be objections from the student body relating to the additional burden of the assessment without a full perception of its need in the context of their own career development.³

However, as Tim MacNamara pointed out during the working party, this issue was way down the line and it was concluded that:

the compulsory nature of the assessment should be re-considered at a later stage in the development of the GSLPA project: but that the end-users of the assessment, the employers of graduates in Hong Kong, would be the major inspiration for the adoption of the machinery that would put the assessment in place.

³The working party wondered whether the assessment would take into account the number of non-native Cantonese speaking undergraduates at the various institutions.

As we thought through this, it became clearer that the involvement of Hong Kong's Captains of Industry in this exercise will need to be great and probably take place at all stages of the project development, in particular when we consider the nature of the communication to be tested. (See issue 2, above.)

In addition to student objections to additional assessment, it was noted that the English Language Teaching operations in the seven tertiary institutions varied considerably. There were variations in the length of the ELT offered, variations in the number of students taught and in the subject disciplines targeted for ELT. In many cases English tuition for students ceased after the first academic year. The students sitting for the first administration of the assessment in 1997 would be that cohort who started their tertiary education in October this year. The majority of these students will have ceased their English classes by 1995 - two years before the assessment might be in place. It was therefore concluded that:

if the present EFL operating systems and practices continue in the various UPGC institutions with no acknowledgement of the fact that language proficiency assessment may be in place by the time the present cohort of undergraduates graduate, then the assessment would:

- a) *disadvantage those students who had not had EFL courses up to the end of their third year of study.*
- b) *make the construction of the assessment impossibly complicated and cumbersome taking into account the varying nature of the courses taken by students.*
- c) *not give future employers of Hong Kong graduates an accurate description of students' language proficiency.*

This is not simply an oblique reference for more UPGC funds: it has become increasingly apparent in Hong Kong that throwing money at teachers doesn't necessarily help to improve students' language skills. It raises another issue - that of more fully integrating language teaching operations in the universities with the main objectives of academic programmes. It seems at present that English for Academic Purposes and the "Business English" courses are often tagged on at the end of a degree course document and English teachers are required to perform miracles in one hour per week. A stronger motivation from within the institutions is maybe required so that, for example, there is at least timetable space for us to work in.

This "issue" has certain input to the next consideration which The Working Party noted, this being the effect that assessment of language proficiency would have on the nature of EFL courses at the various institutions, particularly as the

proficiency would be related to "language for business and commerce". The wash-back effect would necessitate a "teach-to-the-test" practice in class rooms with a proliferation of "practice test" papers and books. While not necessarily a negative educational philosophy, this "test-driven" approach to language teaching could effectively deny the current degree of institutional autonomy in EFL course design, particularly where EFL courses are designed specifically to support the English language needs of specific degree courses

GSLPA will have to be designed with an eye to current teaching practice and course objectives within each institution's language teaching operation.

This point is taken up later in this paper as a problem that relates to the management and administration of GSLPA in Hong Kong.

The Working Party noted that, in an ideal world, the assessment as currently proposed, would have a dramatic and positive wash-back effect on the improvement of language proficiency and language teaching practices.

The assessment, when in place and compulsory would, for reasons of accountability to society, require the provision of funds to cover the need for language education for the three years proceeding assessment for each and any cohort of graduating students. This funding would include, as mentioned, buying time and space for English language teaching

The provision for such funds, that would allow three full years of language teaching at all institutions for all undergraduates, would probably not be forthcoming for the current term of the GSLPA Project. If such a test were in place and compulsory in 1997, those students undergoing assessment (the current first year undergraduates) would be disadvantaged, as mentioned above.

The assessment might better be seen as a pilot test up to and possibly after 1997. As such assessment would not be made on all graduating students, initially, the development of assessment might be better targeted at graduating students from perhaps a few institutions, with co-operation in trialling from other institutions. This would give more time for the development of a reliable and acceptable test - given the current disparate EFL provisions, and would give time for the development and funding of language courses that were more precisely related to the aims of the GSLPA.

Three models for consideration

The complexity of the above constraints and the insights we have gathered on the practicality of the project have brought us to a second planning stage from which we hope to move forward quickly. At this point we feel that it is important to gain some form of consensus quickly on the management and administration of GSLPA across seven institutions with different histories, student populations and academic structures. To address this need three models for the management and administration of GSLPA have been devised and examined for practicality given the complex situation into which the assessment would need to fit.

The first, a traditional "public examination" type of test administered annually to all graduating students, really needs no gloss. The well-known characteristics of these types of test formed the basis for the discussions of our first working party and raised the issues as discussed.

The second model started off with a flippant comment by a colleague, but as we hope you will agree once developed is not without some merit. The comment was:

The job market in Hong Kong being what it is, most graduates in Hong Kong have no difficulty getting a job anyway so why bother to assess their language skills?

Well - model two takes this premise at its face value. All students should be tested but not assessed. The samples of texts gathered from students could be stored and only assessed when a prospective employer or the student concerned requests the service.

The "service" of assessment could be charged for and take either the form of a simple grading (cheap) or a consultation based on the sample of texts that have been gathered (expensive). Alternatively, the samples of texts could be made available to prospective employers for them to make up their own minds on whether the potential employee has the language skills that are required (cheap). This last alternative has the same rationale that many professions adopt through requiring practitioners in a particular field to maintain a portfolio of their work. What started off as a rather light-hearted idea, as you can see, taken a little further does actually start to coalesce into a coherent model

The third model does away with the idea of a monolithic annual test sat by all students in their final year and substitutes for it a series of final year assessments made locally in each institution.

The task of standardising these assessments would be carried out by a proactive group of language assessment specialists drawn from the participating institutions and responsible for the following:

1. the initial provision to participating institutions of: assessment specifications, exemplars of item types, criterion texts, exemplar marking schemes, descriptors for reporting GSLPA
2. external validation of: the tasks set students to elicit performance upon which GSLPA assessment is made; the criterion texts used in assessment; marking schemes
3. external moderation of assessments.

A monolithic, exam-based model of GSLPA would probably at best generate 30 minutes of individual student Oral Interaction and one hour of Writing on which to base a description of communicative proficiency. This alternative model would be more than preferable if it only generated double the quantity of student performance on which an assessment could be made.

1997 and beyond

The final issue which perhaps underlies all the above is, of course, what happens after 1997?

Already the development of English language proficiency testing is taking place in the People's Republic of China with the help of international consultancies mainly from the UK. In particular the University of Guangzhou with the University of Cambridge Local Examinations Syndicate (UCLES) and the London Chamber of Commerce (LCC) are developing a test of Business English for undergraduates in Guangdong Province.

It seems sensible that the GSLPA project team should collaborate with the Chinese test developers and in as far as it is possible or necessary with UCLES and LCC, so that the aims and objectives of proficiency testing on both sides of the border are similar, such that when there is no border the GSLPA may be a suitable system for the assessment of language proficiency for graduating students from universities in southern China including those from the Special Administrative Region of Hong Kong.

So what happens next?

1. Consensus on the form of the management and administration of GSLPA within Hong Kong needs to be attained.
2. The approval and active participation of the local business community needs to be sought to generate the impetus necessary within the community of higher education decision makers, the student body and the community in general; so that this most important project bears the fruit for which it was seeded - dramatic and pervasive change in the university classroom that aims at placing multi-lingual communication skills at the heart of all higher education programmes.

References

- Andrews, S. 1995. Washback or washout? The relationship between examination reform and curriculum innovation. *This volume*.
- Bachman, L.F. 1990. *Fundamental Considerations in Language Testing*. Cambridge. Cambridge University Press.
- Bachman, L.F. & Palmer, A.S. Forthcoming. *Language Testing in Practice*. Cambridge. Cambridge University Press.
- Biggs, J.B. 1995. *Schooling, Learning and TOC*. Paper given at the Opening Ceremony of the INSTEP Workshops for TOC. University of Hong Kong
- Boyle, J. & Falvey, P. 1994. *English Language Testing in Hong Kong*. Hong Kong. Chinese University Press.
- Brindley, G. 1990 (ed.) *The Second Language Curriculum in Action*. Sydney: NCELTR
- Brindley, G. 1995 (ed.) *Language Assessment in Action*. Sydney: NCELTR
- Clark, J.L., Scarino, A and Brownell, J.A. 1994 *Improving the Quality of Learning: A Framework for Target-oriented Curriculum Renewal in Hong Kong*. Hong Kong Language Development Fund/Institute of Language in Education.

- Cram, B. 1995. Self assessment: from theory to practice. Developing a workshop guide for teachers. In Brindley, G. (ed.) 1995.
- Education Commission Report No. 4. 1989. *The Curriculum and Behavioural Problems in Schools*. Government Printer: Hong Kong
- Lam, H.P. 1995. Methodology washback - an insider's view. *This volume*.
- Lewis, J. 1990. Self-assessment in the classroom: A case study. In G. Brindley (ed.) 1990.
- King, R. 1994. Historical survey of English language testing in Hong Kong. In: Boyle, J. & Falvey, P. (eds.) 1994
- King, R. 1995. Opening remarks. *International conference on Testing and Evaluation in Second Language Education*, Hong Kong University of Science and Technology, June 1995.
- Morris, P. 1989. *Curriculum Development in Hong Kong*. Faculty Papers # 7. Faculty of Education: The University of Hong Kong
- Morris, P. 1995. *The Hong Kong School Curriculum: Developments, Issues and Policies*. Hong Kong University Press: Hong Kong.
- Nunan, D. 1988. *The Learner-Centred Curriculum*. Cambridge. Cambridge University Press
- Nunan, D. 1995. Closing the gap between instruction and learning. *TESOL Quarterly*, Vol. 29 # 1 (pp. 133-158)
- Oskarsson, M. 1978. *Approaches to Self-Assessment in Foreign Language Learning*. Oxford: Pergamon Press.
- Rolfe, T. 1990. Self and peer-assessment in the ESL curriculum. In Brindley, G. (ed.) 1990.
- Sengupta, S. 1989. *A study of English teachers' attitudes and expectations concerning Target Related Assessment and students' needs*. Unpublished M.Ed Thesis. The University of Hong Kong.
- Shohamy, E. 1992. The power of tests: A study of the impact of three language tests on teaching and learning. Paper presented at the *14th Language Testing Research Colloquium*. Vancouver, Canada, March 1992.

- Spolsky, B. 1994. The examination-classroom backwash cycle - some historical cases. Paper presented at the *International Language in Education Conference '94*, The University of Hong Kong. Dec.1994.
- Wilkes, M. 1995. Learning pathways and the assessment process: A replication of a study on self- and peer-assessment in the ESL classroom.

4

THE EXAMINATION-CLASSROOM BACKWASH CYCLE: SOME HISTORICAL CASES

Bernard Spolsky
Bar-Ilan University, Israel

The NED defines backwash as "the motion of a receding wave" and notes a metaphorical use "the tremendous backwash of popular enthusiasm" as early as 1864. The Random House Dictionary gives a definition of the use we are concerned with here: "an unexpected and unusually undesirable, subsidiary result or reaction." Strictly speaking, then, the term is better applied only to accidental side-effects of examinations, and not to those effects intended when the *first* purpose of the examination is control of the curriculum. I can find no dictionary support for the language testers' coinage, washback.

In China during the Han dynasty (201 BCE to 8 CE), examinations in which candidates wrote on classical Confucian doctrine replaced patronage as a method of selecting civil servants. To avoid corruption, all essays were marked anonymously, and the Emperor personally supervised the final stage. The system was modified and survived until the end of the nineteenth century. Those who sought after these appointments prepared themselves for these examinations, which became the syllabus for their studies. Thus, though the goal of the examination was to select civil servants, its backwash effect was to establish and control an educational program, as prospective mandarins set out to prepare themselves for the examination.

The first Western scholar to give a detailed (and admiring) account of the Chinese examinations was the late sixteenth century Jesuit explorer and missionary, Matteo Ricci (published in English in 1942). In the Catholic schools of eighteenth century France, the examination was transformed into the ideal method of controlling the teaching in the schools. Thus, backwash became bow or even shock wave. The rules for Christian schools laid down then by Jean-Baptiste de la Salle (1838, 1935) transformed the classroom into a highly organised place, with precise ordering and arrangement of every word and gesture for teacher and pupil, and close factory-like control of all instruction monitored by monthly examinations conducted by teacher and principal.

The justification for using examinations for external control of the classroom is clear, but the inevitable outcome in narrowing the educational process has been repeated over and over again. Once the content of an examination has been bruited abroad, it becomes a more or less precise specification of what

knowledge or behaviour will be rewarded (or will avoid punishment). No reasonable teacher will do other than focus his or her pupils' efforts on the specific items that are to be tested; no bright pupil will want to spend time on anything but preparation for what is to be on the examination. The control of the instructional process then is transferred from those most immediately concerned (the teacher and the pupils) to the examination itself. This may be appropriate when the goal is the rote learning of a specified body of material - a sacred text, a spelling list, a catechism, a multiplication table - but it is constraining and rigid when it pertains to the less defined and more creative aspects of a curriculum. The greater the uniformity, the more the danger of crystallisation and stultification.

The system of examinations in France spread from the Catholic schools to secular institutions and gained in significance. After the French Revolution, examinations continued to grow in importance and were seen at first not as methods of controlling the masses but as a way of allowing access of the educated elite to power. Under Napoleon, however, examinations became the method of controlling a centralised educational system. From the egalitarian service of a revolutionary state to the rigid control of an imperialist one was a short step.

In England, too, the university examination system spread into public life in the nineteenth century, becoming 'a major tool for social policy' by its end. Originally, those taking degrees at the universities (a small proportion of students) had demonstrated their worthiness in public Disputations conducted in Latin. As the Reformation destroyed the oral teaching of Latin, the Disputations lost their value, and were replaced, by written examinations (Latham, 1877). The most important of these examinations was the Cambridge Senate House Examination, later called the Tripos, mainly concerned with mathematics.

By the middle of the nineteenth century, it was widely appreciated that those who had achieved outstanding results on the Cambridge and on similar Oxford examinations were likely to gain high status in later life as leaders of the professions. In a speech given in the House of Commons on 10 July 1833 on the East India Company Bill, Thomas Macaulay made this his argument for using competitive examinations to select cadets for the Indian senior civil service. Twenty years later, Macaulay made the same case in his speech at the second reading of the India Bill on 24 June 1853. He listed men of distinction who had already made their mark by being 'first in the competition of the schools,' among them governors-general of India, lawyers and judges. Even the Leader of the Opposition, Lord Stanley agreed to 'a principle unknown in this country, but which was said to prevail in China, and therefore it might be called the Chinese principle, namely, that of unlimited intellectual competition for

admission to civil offices' although he noted that 'there would also be a practical difficulty in submitting to a proper examination such an enormous and unwieldy number of persons.' In spite of some doubts about making decisions about a man's future on the basis of 'the precocious efforts of youth', the act passed, but it took five years before the system was in place. The first Indian Civil Service examinations were held in 1858, with 67 candidates, twenty-one of whom were selected for cadetships.

The competitive examination established in this way was a democratic but elitist tool, making it possible, superficially at least, to replace patronage with an objective selection procedure in which ability and not birth or connections would determine the result. It was the primary technology employed in the slow transformation from aristocracy to meritocracy, the start of the establishment of a mandarin class in Prussia and England and France made up of brilliant and liberally educated senior civil servants. Their very prominence and public success brought distinction to the technology of tests and examinations, and encouraged in turn their wider use and acceptance. The Local Examinations offered by Cambridge and Oxford and other university boards provided a method of honouring the academic prowess of the more successful elite secondary schools. It served also to co-ordinate the curriculum of the various schools. The lack of centralisation and the fact that schools could choose which Local Board should examine their pupils avoided the uniformity that reigned in France. However, with the examination sovereign, curricular control was taken out of the school.

There was serious criticism of the dangers of competition, the confining of education, and the inappropriateness of using an honours system to measure average students. One early critic was Henry Latham (1877), whose book was 'an enquiry at to what to effect by Examinations and how far we can succeed' (iii). He was convinced that 'within the last thirty years, the agency of Examinations has worked a revolution in the whole province of Education.' (Latham, 1877:1) He characterised examinations as an 'encroaching power' that was influencing education, blurring distinctions between liberal and technical education, narrowing the range of learning through forcing students to prepare by studying with crammers and in cramming schools. Teaching in England was becoming (just as it was, he said, in France) subordinate to examinations rather than its master.

Latham was aware too of some of the problems of technical reliability that were to become a central issue in the debate on objectivity. He noted the difference between examinations that were scored 'by impression' and those that were scored 'by marks' (Latham, 1877:472), and drew attention to the existence of different standards in marks of different examiners. Latham's complaining

voice was the first in a long line of critics who were to bewail, and by their complaints, corroborate the new power of examinations in England at the end of the nineteenth century.

By the end of the nineteenth century, then, the written examination in its traditional form had become stable and entrenched in Europe and the US. It served three main purposes. First, within an instructional context, it could provide a 'stimulant', a 'spur to better performance' (Buck, 1964:7). Second, as the 'encroaching power,' to use Latham's (1877) term, it permitted any external body - a university or a governmental agency - to exert control over the internal operations of educational systems that were becoming increasingly complex. Third, as a 'species of sortition,' to use Edgeworth's (1888) phrase, it was arguably preferable to lotteries or patronage in the selection of civil servants or in the certification of professional qualification.

Examinations were firmly fixed in England, France, Prussia and in the United States as essential instruments in the control of education and in the certification of qualification for employment and further education. Technologically, they had moved from oral to written form, with the open-ended essay and answer the customary item. There were critics of these purposes, who argued that they led to undue pressure and cramming and over-narrow education, and in both countries there was resistance to centralised examination bodies that would have set too rigid a straitjacket on local education

These same debates continued for the next hundred years, but a more momentous attack came not on the *purpose* but on the *technology* itself, when it was demonstrated that the marking of essays and open-ended questions could never reach complete internal consistency or inter-rater reliability. It was this lack of technical reliability that came to worry the scientists who established the discipline of measurement.

How could one avoid unfair manipulation of the system, critics asked, that would favour certain candidates over others? The first published consideration of the effect of error on examinations was a paper read by Professor F. Y. Edgeworth to the Royal Statistical Society in 1888, in which he argued that the general opinion that an examination was only a rough test of merit could be made more precise by application of a part of probability theory, the theory of errors. Physicists, he pointed out, had already demonstrated the existence of error in the measuring of time, distance and weight. Any series of measures had been shown to deviate from the correct measure in a regular fashion, so that small deviations were more common than larger ones. He went on: 'Now there is reason to believe that a similar grouping of divergent estimates prevails when

we are weighing - not physical mass - but intellectual worth.' (Edgeworth, 1888:601) One might expect the marks given a Latin prose by different examiners to deviate in the same way, clustering around a mid-point. In physics, it was possible to check the correct weight with an atomic balance, but in the 'intellectual ponderation' of examining, one must rely on a Standard of Taste, so that the 'mean judgement of competent critics' was the true score, and any deviations were errors.

The truth emerged from these errors: 'By the cumulation of erroneous observations, it is possible to approximate to the truth.' (Edgeworth, 1888:602) Variations in a typical examination resulting from the fluctuating health of the candidate or the special suitability of the questions were not his first concern; the variation that worried him resided in the examiners' readings. Even allowing for the control provided by chief examiners, he assumed an error of 5%, citing in a footnote a number of examples he had seen of much larger variation. As a result, he believed, a proportion of candidates was awarded Honours by chance. He had no simple solution to offer to this problem, but concluded that a public examination is 'a sort of lottery' in which 'the chances are better for the more deserving... It is a species of sortition infinitely preferable to the ancient method of casting lots for honours and offices.' (Edgeworth, 1888:626).

There was no direct response to Edgeworth's challenge. In spite of this, his work had to be repeated regularly in the next hundred years, and still failed to convince many examining bodies. For those who were concerned, however, the issue of reliability raised serious doubts about the fairness of the all-powerful examination. It was this fact that prepared the way for the objective test that was to be presented as a seemingly effective and for a time triumphant solution to the challenge of 'unavoidable uncertainty' that Edgeworth had identified.

The invention of this major technological innovation was a product of related attempts to measure human cognitive ability. The growth of the mental test, as it was called, had begun at least a decade before Edgeworth's papers, but in a paradigm that he was akin to, with the work of Francis Galton, who assumed that anything could (and probably should) be measured. Galton's primary objective was a technique of measuring human attributes that would overcome the limitations on human powers of discrimination. A cousin of Darwin, Galton was the founder of modern statistics and the coiner of the term 'eugenics.' It was an American scholar, James Cattell who studied in Europe and with Galton., who showed how they could be transferred to the educational arena. Cattell (1905) made the transition from theory to practice by showing the relevance of developments in psychology to the determination of individual differences and the selection of individuals. His motivation was strongly social, and evinced

recognition of the social effects of unfair or inaccurate testing His cry is one that is still valid today:

It is quite possible that the assigning of grades to school children and college students as a kind of reward or punishment is useless or of worth; its value could and should be determined. But when students are excluded from college because they do not secure a certain grade in a written examination, or when candidates for positions in the government service are selected as the result of a written examination, we assume a serious responsibility. The least we can do is to make a scientific study of our methods and results. (376)

Cattell's work in the United States added to that of other psychologists like Thorndike was slowly introducing objective measurement into the examination business, but the major breakthrough in testing came from the work of a French doctor, Binet, with a pragmatic need to deal with a real problem. The main purpose of Binet's work was to deal with mental retardation, of which he believed only a multiplicity of tests would give a clear picture. When reports of Binet's work first reached the U.S. in 1906, they made little impact. However, after a visit to Europe in 1908, Henry Goddard, psychologist at the New Jersey Training School for Feeble-minded Boys and Girls in Vineland, decided to try the new scales and was so impressed with the results that he immediately published an outline, and was successful in 1910 in persuading the physician-dominated American Association for the Study of the Feeble-minded to adopt the Binet tests as the way to diagnose mental subnormality (Zenderland, 1987:46).

The spread of the Binet scale was furthered by the work of Terman, who, in 1910, when he received an appointment to the psychology department at Stanford University, began work to revise it. The mental tests, he believed, would save money by permitting more focused teaching for the retarded and so reduce crime and the cost of prisons. In England, the Binet- Simon scale was introduced to widespread use by Cyril Burt (1921) who, in the years immediately after the first world war, was psychologist for the education department of the London County Council. But its wide public acceptance came in the United States after the first World War, thanks largely to the organising and public relations work of Robert Yerkes, who firmly established mental measurement with the American public.

The modern objective test emerged in the context of the continuing tension between the unmistakably effective and awesome power of the examination and

the manifest uncertainty of its scores. The technological innovation was accepted reluctantly at first by only a few educators, but rapidly became the dominant force in American institutional testing. The claims about the use of intelligence tests in America in the first World War gave objective tests a major boost. The Army Alpha tests, with all their weaknesses and confusion, were the beginning of mass objective testing of mental abilities and were to be sold to the general public as the solution to the unreliability of traditional examinations. For Yerkes, man was just as measurable as machine.

In some quarters, there was strong criticism of the new intelligence tests, especially as their racist implications emerged. Brigham (1923) who had worked with Yerkes emphasised the racial implications. 'We must now frankly admit the undesirable results which would ensue from a cross between the Nordic in this country with the Alpine Slav, with the degenerated hybrid Mediterranean, or with the Negro, or from the promiscuous intermingling of all four types.' (1923:208) American intelligence, he concluded, would continue to decline as long as immigration was free: 'Immigration should not only be restrictive but highly selective' . (1923: 210). The journalist Walter Lippmann was one of the most outspoken of the opponents (Block and Dworkin, 1976), but even he could see some redeeming value in the new tests. By 1920, most state universities were offering courses in educational measurement, and by 1923 half the business of the Teachers College Bureau of Publications was in test and scales, with over 12,000,000 copies sold in the previous ten years (Joncich, 1968:389). Test publishing was becoming big business.

The movement gained more power when Brigham and his psychometrically inclined colleagues became involved in the growth of centralised college admission in the US. While he was a teacher at Princeton in the 1920s, Brigham had begun to experiment with objective intelligence tests that he gave to incoming freshmen. Under his influence, Princeton began to require the test of all candidates for admission in 1925; in the same year, the College Board commissioned Brigham and other university psychologists to start constructing the tests that became the Scholastic Aptitude Test (Whitla, 1964:57).

The College Entrance Examination Board (its full title) had been conceived during the last decades of the nineteenth century to remedy the lack of co-ordination or consensus among colleges in requirements or methods of selecting students for admission. Formally constituted in November 1900 by the founding group of a dozen colleges (Farrand, 1926), by 1925, 35 colleges had joined the Board and there were twenty examining groups, 316 centres, 20,000 candidates, and 600 readers. At first, the Board had planned to leave the specifications and requirements for the examinations and the resulting power over the school curriculum to authorities in each field of study, such as the

Modern Language Association for English and languages. However, as these bodies were more interested in scholarship than in the teaching of their subject, in 1907 the Board appointed its own commissions of specialists in each of the fields. The early philosophy of the Board was expressed by Lowell (1926). Examinations, he believed, should have three purposes: to measure progress, to control instruction, and to set a standard. The Board saw control of the curriculum and of the pupils as its key mission.

Quite different arguments started to be heard when Brigham joined the Board as assistant secretary. Carl Brigham chaired the committee on the Scholastic Aptitude Test. In a report (College Entrance Examination Board, 1926:44-63), he explained the psychometric principles behind the use of tests as predictors of college success. But these tests did not provide perfect prediction. Part of the problem resided in the tests, which did not (and could not) measure all the relevant factors with perfect accuracy, and part on the difficulty of measuring the outcome, college success. The committee stressed the need for multiple measures, knowing that any single measure could be wrong.

There was opposition in America to the psychometric movement, especially from those who hoped to use tests and examinations as a means of controlling the school curriculum. Edgeworth's papers, brilliant though they were, and accepted by psychologists like Burt in Britain and Brigham in the U.S., did not immediately and everywhere carry the day. In much the same way that there would be objection to a suggestion to drop diving and gymnastics and other judged events from the Olympic games, so demands continued to be made for persisting with the use of the essay that had become the archetype of the nineteenth century English examination.

The battle was hotly waged inside the major testing institutions. In a democratic gesture mitigating their power, the College Entrance Examination Board had always recognised the need not to 'hamper the independence and initiative of individual teachers' by allowing its examinations to become too firmly set. They also wanted to permit a candidate to show understanding and power of expression, and not just 'test his ability to reproduce in a more or less mechanical manner bits of information imparted to him by a drillmaster.' To this end, they avoided precise prescription of requirements or of the form of examination and regularly changed the membership of examining committees 'to facilitate variations in the form of its examinations and to give suitable recognition to the different shades of opinion that exist among American scholars and teachers.'

Given these quite liberal and humanistic view of the examiners, one can readily imagine the uphill battle that Carl Brigham, its new associate secretary,

faced in trying to convince the Board of new psychometric truths. In the 1933 Report (College Entrance Examination Board, 1933), he presented the objectivity versus control issue and drew the Board's attention to the two 'partially incompatible' purposes of the Board, the one to control institutions and the other to measure individuals. The Board's 'definitions of requirements,' which had been questioned by teachers as interfering with the schools' freedom, could be written more generally, but they would still 'collide with the nature and composition of the examination papers.' Brigham reminded the Board of the unreliability of measuring devices. Testers had long been aware of this problem, and so they thought of a grade not as a fixed point but rather as 'one particular grade drawn by chance from a hat which contains all of that candidate's grades obtained from an infinite series of examinations set and read to measure exactly that same trait.' Unlike the colleges which acted as though a reported grade were an exact and fixed number, testers saw it as surrounded by its standard error, 'which would include two-thirds of the grades above and below it.' The purpose of the 'new-type' testing movement was essentially to reduce this error and so to rate an individual more fairly. To testers with this aim, concern for control of the curriculum would be 'heretical.'

The Board's Examiners, however, considered 'that their duty is to exercise a benevolent control over the curriculum.' From this point of view, questions should be included in the examination simply because schools must train students to answer them. But such questions, Brigham said, were often quite unsuitable to 'describe an individual with a respectable minimum of variation in rating.' Brigham thought the underlying concern of the objective testing movement was reasonable, but 'the crusader spirit has converted good devices for measuring into a prescription of what should be measured.' Thus, some testers had come to assume that whatever they could measure was what should be taught.

By the early 1930s, the battle of reliable testing of individuals versus using examinations to control the educational system was virtually over, with the triumph of Brigham and the psychometrists. Curricular control became backwash, an unwanted effect, as teachers and commercial cramming schools saw it as a challenge to beat the objective test on its own ground, by preparing students to answer multiple choice questions.

I think we come here to the crux of the backwash problem. There may be fields in which the specification of goals for a curriculum can be no more than a list of items to be memorised. In such a field, a sampling from this universe through a carefully devised objective test might serve as a satisfactory measure of individual achievement, and a listing of the universe could serve as the guideline for a satisfactory syllabus.

In the case of language proficiency, however, it has long been clear that there is a paradox: the integrative skills do not lend themselves to easily reliable measurement, and the discrete items that can be objectively and reliably measured are probably not a valid representation of the desired proficiency, and are certainly not the basis for a satisfactory curriculum. The traditional examination provided the basis for a good curriculum, but a poor basis for individual proficiency testing (as witness the lack of reliability found in the UCLES CPE in recent studies), while the objective test provides a good basis for individual testing until cramming schools convert it into a tedious and vapid curriculum.

The paradox might be illustrated with vocabulary tests. It is clear from many studies that the best and quickest measure of overall language proficiency is a vocabulary test; the better you know a language, the more words you know. But it is equally clear that once vocabulary testing were established as the only method of assessing competence, the backwash to instruction would lead not just to boring and ineffective teaching (classes called on to memorise long lists of words, as François Gouin once did in a fruitless attempt to master German) but also would vitiate the effect of the testing, which would become a test not of language proficiency but of vocabulary learning.

How serious this is in practice must of course depend on the attractiveness of the examination pass: one suspects that a good proportion of the Asian students who take TOEFL every year, constituting over 70% of the annual 600,000 plus candidates, were more highly motivated to pass the examination than to learn English.

What I have been arguing here must perforce remain a plausible hypothesis until there is more detailed research of the type initiated by Alderson and Shohamy on the effect of backwash, or of the type undertaken by Berwick to show the weakness of TOEFL, in part presumably because of the tendency to cramming, as a predictor for many Asian students. Like gunpowder, another Chinese invention, the examination is far from an unmixed blessing, capable of dangerous misuse as well as useful for good ends. One of the unfortunate effects of the technological revolution of psychometrics was to divert attention from the use of the test to its use; the new ethically responsible concern for validity of tests and their uses offers an opportunity to once again harness backwash.

Bibliography

- Binet, A. and T. Simon 1916. *The development of intelligence in children*. Baltimore, Williams and Wilkins.
- Block, N. J. and G. Dworkin (ed.) 1976. *The IQ controversy: critical readings*. New York, Pantheon Books.
- Brigham, C.C. 1923. *A study of American intelligence*. Princeton, Princeton University Press.
- Buck, P. H. 1964. 'Examinations: a retrospective view at Harvard' in L. Branson (ed.): *Examining at Harvard College*. Boston, Committee on Educational Policy, Harvard University. 4-37.
- Burt, C. L. 1921. *Mental and scholastic tests*. London, London County Council.
- Cattell, J. M. 1973 (1947). *Man of Science*. New York, Arno Press.
- College Entrance Examination Board 1926. *The work of the College Entrance Examination Board: 1901-1925*. Boston, Ginn and Company.
- College Entrance Examination Board 1933. *Thirty-third Annual Report of the Secretary*. New York, College Entrance Examination Board.
- Edgeworth, F. Y. 1888. 'The statistics of examinations.' *Journal of the Royal Statistical Society* 51: 599-635.
- Farrand, W. 1926. 'A brief history of the College Entrance Examination Board' in College Entrance Examination Board (ed.): *The work of the College Entrance Examination Board: 1901-1925*. Boston, Ginn and Company. 21-30.
- Galton, F. 1883. *Inquiries into human faculty and development*. New York, MacMillan and Co.
- Joncich, G. M. 1968. *The sane positivist: a biography of Edward L. Thorndike*. Middletown, Connecticut, Wesleyan University Press.
- La Salle, J.-B. 1935. *The conduct of the schools*. New York, McGraw Hill Book Company.
- La Salle, J.-B. d. 1838. *Conduite des Ecoles Chrétiennes*. Paris, J. Moronval.

- Latham, H. 1877. *On the action of examinations considered as a means of selection*. Cambridge, Deighton, Bell and Company.
- Macaulay, T. B. 1853. *Speeches, parliamentary and miscellaneous*. London, Henry Vizetelly.
- Macaulay, T. B. M., Baron 1899. *The Works of Lord Macaulay*. Boston, Houghton Mifflin.
- Ricci, M. 1942. *China in the sixteenth century. the journals of Matthew Ricci, 1583-1610*. New York, Random House.
- Thorndike, E. L. 1904. *An introduction to the theory of mental and social measurements*. New York, The Science Press.
- Thorndike, E. L. 1910. Handwriting. *Teachers College Record* 11(2).
- Whitla, D. K. 1964. Uses of the multiple choice test. In: L. Bramson (ed.): *Examining in Harvard College*. Cambridge, Massachusetts, Committee on Educational Policy, Harvard University. 56-63.
- Yerkes, R. M. 1911. *Introduction to psychology*. New York, Henry Holt and Company.
- Yerkes, R. M. (ed.) 1921. *Psychological examining in the United States Army*. Washington, Government Printing Office.
- Yerkes, R. M., J. W. Bridges, and R. S. Hardwick 1915. *A point scale for measuring mental ability*. Baltimore, Warwick and York, Inc.
- Zenderland, L. 1987. The debate over diagnosis: Henry Herbert Godard and the medical acceptance of mental testing. In: M. M Sokal (ed.): *Psychological testing and American society 1890-1930*. New Brunswick, Rutgers University Press. 46-74.

5

WASHBACK OR WASHOUT ? THE RELATIONSHIP BETWEEN EXAMINATION REFORM AND CURRICULUM INNOVATION

Stephen Andrews

Dept. of Curriculum Studies, University of Hong Kong

Introduction

The influence of examinations on teaching and learning has been a frequent topic for discussion among educators for many years (see, for example, Wiseman 1961, Davies 1968, Kellaghan et al 1982, Oxenham 1984, Swain 1985, Heyneman 1987, Morris 1990, Wall and Alderson 1992, Alderson and Wall 1993). In the general education literature, the favoured term to describe this phenomenon is 'backwash', while in language education there seems to be a preference for 'washback'. In recent years, it has become increasingly common to suggest that washback can be utilised as a 'lever for change' (Pearson 1988) in the curriculum. It is less than clear, however, whether what happens in classrooms in response to examination innovation corresponds very closely to the effect intended by the innovators. This paper reports on a small-scale study which attempts to investigate the relationship between washback and curriculum innovation.

Washback

Washback - defined by Hughes as "...the effect of testing upon teaching and learning" (Hughes 1989:1) - is an area where the literature in both language education and general education is long on assertion and short on empirical study.

There seems to be consensus that there is such a thing as a washback phenomenon (see, for example, Pearson 1988 and Morris 1990). Thus it is not unusual to find statements like the following : "It has frequently been noted that teachers will teach to a test : that is, if they know the content of a test and/or the format of a test, they will teach their students accordingly" (Swain 1985:43) and "In many education systems the key question for students, teachers, parents, school administrators, and even inspectors is not, 'Are students gaining in communicative competence ?' but, 'Are they on course for the examination ?' " (Johnson 1989:6).

At the same time, by contrast, there is a relative lack of empirical evidence to support or throw light on such claims. Hughes' 1988 study is an exception in this respect : his findings indeed appear to confirm the existence of some kind of washback phenomenon (Hughes 1988). As Alderson and Wall comment, however, his study does not offer any insights into the nature of washback, its precise effects or how it works (Alderson and Wall 1993:125-126).

Indeed, as Alderson and Wall point out, washback is both a complex and an ill-defined phenomenon (1993:117). As an illustration of the imprecision which is apparent in much discussion of washback, they present 15 different versions of washback hypothesis, among them the following :

A test will influence teaching
A test will influence learning
A test will influence what teachers teach
A test will influence how teachers teach
A test will influence what learners learn
A test will influence how learners learn
(Alderson and Wall 1993:120-121).

Washback has often been seen as having a harmful effect upon educational practices. Oxenham sums up this view as follows : "The harm of centralised examinations is said to spring from the restrictions they will impose upon curricula, teachers and students....Their almost inescapable bias is to encourage the most mechanical, boring and debilitating forms of teaching and learning" (Oxenham 1984:113). Recently, however, there has been a growing tendency to explore the possibilities of using washback to promote pedagogical ends, of "working for washback", as Swain describes it (Swain 1985). Brooke and Oxenham, for instance, recommend that "...a prime line of reform should bend examinations to fit the wider objectives of education. By that kind of judo trick, examinations....should promote and reward better teaching and learning" (Brooke and Oxenham 1984:175).

In spite of such recommendations and the fact that there have been a number of assessment innovations which have been introduced with pedagogical ends in mind (see, for example, Swain 1985 on the development of a test of French, UCLES 1992 on the PISSET project), there have been very few research studies which throw light on the actual classroom impact of assessment changes. One exception to this is Wall and Alderson (1992), in which they discuss their Sri Lanka study, investigating the classroom impact of changes to the Sri Lanka O-level English exam. One of the more illuminating findings of the study is the following :

"Although the examination has had a demonstrable impact on the content of language lessons, it has had virtually no impact on the way that teachers teach English....[This] lack of impact may be because of a lack of understanding on the part of teachers of what might be an appropriate way to prepare students for the examination. It may even be that the exam itself - and this may indeed be true of all exams - does not and cannot determine how teachers teach, however much it might influence what they teach" (Alderson and Wall 1993:127).

Background to the Present Study

In Hong Kong, in the field of language education, there have been a number of reforms to the public examinations which have deliberately sought to initiate or reinforce changes in classroom practice (see the discussion in King 1994 for examples). The Use of English (UE) examination is a particular case in point.

The UE is taken by about 20,000 Secondary 7 (13th year) students in Hong Kong each year. The results are used as admissions criteria by tertiary institutions and are also used by many employers as a measure of the English language proficiency of prospective employees. The current version of the UE was introduced in 1989, and "...the aim was to produce a positive washback effect on teaching throughout the whole of the sixth-form course" (King 1994:17). For example, the *Practical Skills for Work and Study component* (known as WASPS) was specifically designed to encourage a focus upon the integration of language skills in the classroom.

The 1989 UE had four components (listening, writing, reading and language systems, and WASPS). At the time it was introduced, it had no oral component. Pressure from the universities, however, led the Hong Kong Examinations Authority (HKEA) to review its original decision not to include an oral. There was concern about the level of spoken English among the undergraduate population, and it was felt that there might be some improvement if the UE included an oral component and therefore encouraged continuing attention to the speaking skill in Secondary 6 and 7.

In November 1990, the HKEA set up a Working Party to draw up proposals for an oral component. The Working Party (comprising seven tertiary teachers, two secondary teachers, one British Council teacher, and one HKEA test development officer) was fully aware of the intention that the oral component should have a positive washback effect on classroom practice in Secondary 6 and 7, and a two-part test design was eventually proposed, incorporating both individual oral presentation and group discussion. The new version of the UE, incorporating the

oral component, was administered for the first time in 1994, so that all students entering Secondary 6 since September 1992 have been following a course of English study leading towards the five-component UE.

The following comment from a colleague acted as a catalyst for this small-scale study: "Washback from exams is undoubtedly crucial in curriculum innovation. It is definitely not a simple one to one relationship, however...the intentions of examiners and the strategies teachers adopt to outwit them can be very different (watch it happen in the UE Oral!)" (R.K.Johnson, personal communication). The purpose of this small-scale study is to try to gain some insight into the actual washback effect of the UE Oral, and to explore some of the complexities of the relationship between examination reform and curriculum innovation by shedding light on the intentions of examiners, in this case the UE Oral Working Party, and on the strategies adopted by teachers.

Research Questions

The principal questions being investigated in this study are as follows:

- 1) How do the examiners - those who aim to use exam washback as a catalyst for curriculum innovation ('lever for change') - perceive :
 - a) the relative importance of different possible impacts of the exam upon students, teachers, and what happens in the classroom ?
 - b) the relative priority they would give to different pedagogical strategies which might be used to prepare students for the exam ?
- 2) How do the teachers - receivers/implementers of the curriculum innovation - perceive:
 - a) the actual impacts of the exam upon students, teachers, and what happens in the classroom ?
 - b) the relative priority they would give to different pedagogical strategies which might be used to prepare students for the exam ?
- 3) To what extent is there a match between :
 - a) 1a and 2a ?
 - b) 1b and 2b ?
- 4) To what extent are exam-related published textbooks an influence on the pedagogical strategies teachers employ in preparing students for an exam?

Methodology

Two parallel questionnaires were used to gather data relating to the research questions. Version 1 was submitted to all ten UE Oral Working Party members still in Hong Kong. There was 90% return rate. Version 2 was submitted to a range of Secondary 6 and 7 teachers, all of whom either had attended or were attending workshops on the UE at the University of Hong Kong. 22 respondents were attending the workshops at the time they completed the questionnaire : their responses were given before they had had any input relating to the Oral. 73 respondents had attended workshops between eight months and three years previously. All of this latter group would have had some input concerning the Oral. Of the 95 teacher respondents, 30 had taught Secondary 6 and 7 both before and after the introduction of the UE Oral, while 62 others were currently teaching Secondary 6 or 7, but had not done so before the Oral was introduced.

Results

One of the more interesting sets of findings to emerge from this study concerned the consistency of view among the members of the Working Party. As the responses to the first three questions indicate, there was a high degree of consistency of opinion about the general importance of washback :

[In tables 1 to 3 the figures represent the actual number of respondents - n = 9]

1) When you were involved in the development of the UE Oral, were you aware of its potential washback?	YES 9 / NO 0
2) Did you think that the washback would be positive, or negative ?	POSITIVE 9 / NEGATIVE 0
3) How important did you consider the potential washback effect to be ?	EXTREMELY IMPORTANT 8/ VERY IMPORTANT 1
<i>[rated on a 5-point scale from NOT AT ALL IMPORTANT to EXTREMELY IMPORTANT]</i>	

Table 1

Thereafter, however, when the questions focused on the importance of specific forms of washback and particular classroom practices, there was rather less consistency, as the tables below illustrate. Section C of the questionnaire, for

example, asked how important members of the Working Party considered the potential washback to be in the following areas :

	NOT AT ALL IMPORTANT			EXTREMELY IMPORTANT	
an increase in Forms 6 and 7 in :	1	2	3	4	5
4) students' spoken language proficiency ?			1	4	4
5) students' motivation to speak English ?			1	3	5
6) students' confidence in speaking English ?				4	5
7) students' willingness to spend class-time on the speaking skill ?		1		3	5
8) students' enjoyment of English lessons ?		2	3	3	1
9) students' ability and willingness to work independently of the teacher ?		2	4	3	
10) teachers' interest in developing their students' ability to speak English ?			2	1	6
11) teachers' willingness to spend class-time on the speaking skill ?				2	7
12) teachers' enjoyment of English lessons ?		2	4	1	2
13) the amount of class-time spent on oral activities ?		1		2	6
14) the amount of group-work in English lessons ?			2	3	4
15) the integration of oral activity with work on other skills ?			3	4	2

Table 2

Similar inconsistencies can be seen in the Working Party's answers to Section D, which asked respondents to rate the importance, in preparing students for the UE Oral, of spending time on the following :

	NOT AT ALL IMPORTANT			EXTREMELY IMPORTANT	
	1	2	3	4	5
16) practising the UE Oral format ?		2	1	4	2
17) giving students a framework to follow in making an oral presentation ?	1	4	1	2	1
18) teaching students techniques for summarising ?	2	2	3	2	
19) teaching students techniques for making notes ?	1	3	1	3	1
20) explaining techniques for making an oral presentation ?		4		4	1
21) explaining techniques for participating in group discussion ?		3		3	3
22) practising techniques for making an oral presentation ?		1	1	4	3

23) practising techniques for participating in group discussion ?			1	3	5
24) teaching specific phrases for opening and closing an oral presentation ?	1	5	1	1	1
25) teaching specific phrases for highlighting main points in an oral presentation ?	1	4	3	1	
26) teaching specific phrases for eg interrupting, supporting other speakers in a group discussion ?		1	3	2	3
27) teaching students to use appropriate body language (including eye-contact) in an oral presentation/group discussion ?		1	3	4	1
28) getting students to speak as much English as possible ?				3	6
29) getting students to take part in simultaneous group discussion ?				2	7
30) getting students to speak to each other in English when the lesson focus is on other skills?			2	3	4
31) getting students to perform tasks specifically designed to integrate practice of the skills ?		1	3	4	1
32) getting students to focus on communicating their ideas in English?				5	4
33) getting students to focus on the grammatical accuracy of their spoken English ?		3	5	1	
34) helping students to develop strategies so that when they are speaking English they can make the best use of whatever English they know ?			1	4	4
35) encouraging students to monitor their own and their peers' spoken language proficiency ?		2	2	2	3
36) getting students to relax and to speak English naturally and informally ?			3	2	4
37) getting students to be open to and interested in the ideas of other people ?		1	3	3	2
38) teaching set phrases to fill out any oral presentation or contribution to a group discussion ?	5	1	1	1	1
39) getting students to respond sensitively and appropriately to others in a group discussion ?			1	2	6
40) getting students to participate in such out-of-class activities as debates and speaking competitions ?	1	2	3	1	2

Table 3

Before attempting any comparison between the views of the UE Oral Working Party and the perceptions of those teachers with experience of teaching towards the UE both pre- and post-Oral, the limitations concerning the latter should be noted. There are only thirty respondents in this particular sub-group of the teacher sample, the slant of questions 4 to 15 assumes a positive washback effect (which

may skew the responses somewhat), and the responses are only an indication of perceptions. As a result, only tentative interpretations are possible.

Nevertheless, if one looks at Tables 4 and 5, some interesting comparisons can be made. Both groups emphasise the willingness of teachers to spend time on the speaking skill. However, it is noticeable that the UE Oral Working Party places greater emphasis on the impact upon students, in terms of their confidence and also their proficiency, in addition to their motivation. The teachers' perceptions suggest that there may be acceptance by teachers and students of the need to spend time on the speaking skill, and that there may indeed have been quantitative changes in the amount of time spent on speaking and group-work, but that there has been relatively less impact on students' confidence and proficiency, both of which are absent from the list of those changes perceived to be most salient.

Areas of potential washback considered most important by the UE Oral Working Party

An increase in :		
	<u>Mean</u>	<u>S.D.</u>
1) T. willingness to spend class-time on the speaking skill	4.78	.4157
2) S. confidence in speaking English	4.56	.4969
3) = S. motivation to speak English	4.44	.6849
3) = T. interest in developing their students' ability to speak English	4.44	.8315
3) = the amount of class time spent on oral activities	4.44	.9558
6) = S. spoken language proficiency	4.33	.6667
6) = S. willingness to spend class-time on the speaking skill	4.33	.9428
8) the amount of group-work in English lessons	4.22	.7857

Table 4

**Perceptions of change since introduction of UE Oral
by 30 teachers with experience of F6/7 teaching both pre- and post-UE Oral**

An increase in :		
	<u>Mean</u>	<u>S.D.</u>
1) = T. willingness to spend class-time on the speaking skill	3.77	.7157
1) = S. willingness to spend class-time on the speaking skill	3.77	.7608
3) T. interest in developing their students' ability to speak English	3.69	.7476
4) = S. motivation to speak English	3.67	.8300
4) = the amount of group-work in English lessons	3.67	.6992
6) the amount of class-time spent on oral activities	3.6	.6633
7) the integration of oral activity with work on other skills	3.45	1.0368
8) T. enjoyment of English lessons	3.3	.9363

Table 5

Tables 6 to 9 show how the two groups of respondents perceive the relative importance of various pedagogical strategies in preparing students for the UE Oral. In the report of these results, it should be noted that the teacher group is considerably larger than before, with 92 respondents. When one compares the ratings of the two groups, there are a number of similarities, but also some interesting differences. Among the ten strategies rated most important by the two groups (Tables 6 and 7), there are seven which appear in both lists. It is noticeable, however, that the three strategies appearing only on the teachers' list (those rated 6th, 7th and 10th in importance) are all of the same type : very teacher-centred strategies focusing on specific techniques for the test.

**Perceived priorities in preparing students for the UE Oral
Group A (UE Oral Working Party) : TOP 10**

	<u>Mean</u>	<u>S.D.</u>
1) getting students to take part in simultaneous group discussion	4.78	.4157
2) getting students to speak as much English as possible	4.67	.4714
3) getting students to respond sensitively and appropriately to others in a group discussion	4.56	.6849
4) = getting students to focus on communicating their ideas in English	4.44	.4969

4) = practising techniques for participating in group discussion	4.44	.6849
6) helping students to develop strategies so that when they are speaking English they can make the best use of whatever English they know	4.33	.6667
7) getting students to speak to each other in English when the lesson focus is on other skills	4.22	.7857
8) getting students to relax and to speak English naturally and informally	4.11	.8749
9) practising techniques for making an oral presentation	4.00	.9428
10) teaching specific phrases for eg interrupting, supporting other speakers in a group discussion	3.78	1.0304

Table 6

Perceived priorities in preparing students for the UE Oral
Group B (92 Current F6/7 Teachers) : TOP 10

	<u>Mean</u>	<u>S.D.</u>
1) getting students to speak as much English as possible	4.5	.6803
2) practising techniques for participating in group discussion	4.29	.6812
3) getting students to take part in simultaneous group discussion	4.25	.7317
4) practising techniques for making an oral presentation	4.23	.7094
5) getting students to focus on communicating their ideas in English	4.21	.8902
6) explaining techniques for making an oral presentation	4.17	.6529
7) explaining techniques for participating in group discussion	4.16	.6272
8) getting students to respond sensitively and appropriately to others in a group discussion	4.13	.7692
9) getting students to speak to each other in English when the lesson focus is on other skills	4.06	.8969
10) teaching students techniques for making notes	4.02	.8119

Table 7

If one looks at those strategies rated least important by the two groups of respondents (Tables 8 and 9), there are again interesting similarities, with two strategies common to both lists. Perhaps the most noteworthy difference is the relative lack of enthusiasm among members of the UE Oral Working Party, in comparison with the teachers, for 'frameworks' and teaching set phrases.

**Perceived priorities in preparing students for the UE Oral
Group A (UE Oral Working Party) : BOTTOM 5**

	<u>Mean</u>	<u>S.D.</u>
21) giving students a framework to follow in making an oral presentation	2.78	1.2273
22)=teaching students techniques for summarising	2.56	1.0657
22)=teaching specific phrases for opening and closing an oral presentation	2.56	1.1653
24) teaching specific phrases for highlighting main points in an oral presentation	2.44	.8315
25)teaching set phrases to fill out any oral presentation or contribution to a group discussion	2.11	1.4487

Table 8

**Perceived priorities in preparing students for the UE Oral
Group B (92 Current F6/7 Teachers) : BOTTOM 5**

	<u>Mean</u>	<u>S.D.</u>
21) teaching students techniques for summarising	3.69	.8649
22) encouraging students to monitor their own and their peers' spoken language proficiency	3.58	.8838
23) getting students to participate in such out-of-class activities as debates and speaking competitions	3.57	1.0141
24) teaching set phrases to fill out any oral presentation or contribution to a group discussion	3.32	.8629
25) getting students to focus on the grammatical accuracy of their spoken English	3.28	.8721

Table 9

Some of the most revealing information in the study is to be found in the responses of the teachers to questions about the materials they use in preparing their students for the UE Oral. As the figures in Table 10 indicate, an estimated two-thirds of class-time is spent working with exam-focused published materials, with practice tests seeming to occupy the most time.

When preparing your students for the UE Oral,
what proportion (i.e.%) of your class-time is based on the following ?

a)	Published books of UE Oral practice tests	34.34%
b)	Published books aiming to develop skills for the UE Oral	32.55%
c)	Other published materials (not linked to the UE Oral)	16.22%
d)	Your own materials	16.7%

Table 10

As Table 11 illustrates, those teachers making extensive use of their own materials are very much in a minority. The relative popularity of the published materials mentioned is significant, with the most popular being those which are highly exam-specific, focusing exclusively on the precise skills required for the UE Oral exam and on the exam format itself.

Those claiming to spend 40% or more time using their own materials :

9.78%	
40%	3 teachers
50%	3 teachers
70%	1 teacher
100% [divided between c) and d)]	1 teacher
100%	1 teacher

Published materials mentioned as being used :

Witman Oral 1 and 2	72.83%
Expression	50%
Macmillan UE Practice Papers	16.3%
Speaking Skills (Macmillan)	9.8%
Talking It Through (Longman)	7.6%

Table 11

Conclusions

On the basis of the current study, a number of tentative conclusions can be drawn. For instance, with regard to the possibility of engineering precisely targeted pedagogic change by means of the washback from public examinations, it seems reasonable to make the following comments :

- i) there is an obvious sense in which targeting can be precise : in relation, for example, to a particular skill area, as can be seen with the UE Oral;
- ii) otherwise, however, precise targeting seems to be much more problematic, especially when those responsible for trying to introduce the innovation (in this case, the UE Oral Working Party) have widely diverging views about the exact nature of the change they want to engineer;
- iii) there does seem to be a phenomenon to which one could attach the label washback. As a tool to engineer curriculum innovation, however, washback seems to be a very blunt instrument, one which may have relatively predictable quantitative effects on, for example, the time allocated to different aspects of teaching and on the content of that teaching, but rather less predictable qualitative effects upon the teaching-learning process and what actually takes place in classrooms.

It is in part the mismatch between the views of the UE Oral Working Party and those of the teachers which gives rise to the doubts expressed in the previous sentence. However, their apparently contrasting views of the relative importance of different pedagogical strategies raise as many questions as they answer. In particular, one has to ask :

- i) how far do the responses of the teachers (as suggested by the relative flatness of their ratings) reflect not having thought about pedagogical strategies very much?
- ii) how far do the responses of the teachers reflect thinking which has been conditioned by the textbook being used ?
- iii) how can we really begin to understand washback without looking much more closely (as Alderson and Wall 1993 suggest) at what happens in the classroom?

As for the role which exam-related textbook materials play in the relationship between washback and curriculum innovation, this study highlights the importance of that role, even in relation to a skill such as speaking, which one might imagine as needing to depend relatively little upon materials. In any characterisation of the role played by published materials, which should itself be the subject of further research, it is important to emphasise that although the design of an exam-related textbook may be based on information from the examining body responsible for an innovation (in this case, the HKEA), the final product will not be moulded according to the innovators' view of what is desirable

in terms of teaching, but rather according to the publisher's view of what will sell. In the Hong Kong context, at least, this unfortunately tends to lead to materials which are very examination-specific, and which represent a limiting of focus for teachers and students rather than a broadening of horizons.

In conclusion, it would appear that the present study, rather than clarifying the relationship between examination washback and curriculum innovation, has only served to emphasise its complexity. Washback may be of value as a tool to facilitate curriculum innovation in situations where assessment procedures, textbook-writing, and teacher-training are able to work together in harmony. Where these various elements of a curriculum innovation are not totally in harmony, however, then the successful implementation of an innovation is bound to be in some way compromised.

References

- Alderson, JC and Wall, D (1993) "Does washback exist ?" *Applied Linguistics* 14,2.
- Brooke, N and Oxenham, J (1984) 'The influence of certification and selection on teaching and learning'. In Oxenham, J (ed.) (1984)
- Davies, A (1968) *Language testing symposium - a psycholinguistic approach* Oxford:Oxford University Press.
- Heyneman, SP (1987) 'Uses of examinations in developing countries : selection, research and education sector management' *International Journal of Educational Development* 7,4.
- Hughes, A (1988) 'Introducing a needs-based test of English language proficiency into an English-medium university in Turkey'. In Hughes, A (ed.) *Testing English for university study* ELT Documents 127. Modern English Publications.
- Hughes, A (1989) *Testing for language teachers* Cambridge:Cambridge University Press.
- Kellaghan, T, Madaus, GF and Airasian, PW (1982) *The effects of standardized testing* London:Kluwen, Nijhoff Publishing
- Johnson, RK (1989) *The second language curriculum* Cambridge:Cambridge University Press.

- King, R (1994) 'Historical survey of English language testing in Hong Kong' In Boyle, J and Falvey, P (eds.) *English language testing in Hong Kong* Hong Kong:Chinese University Press.
- Morris, P (1990) *Curriculum development in Hong Kong* Hong Kong:Faculty of Education, University of Hong Kong.
- Oxenham, J (ed.) (1984) *Education versus qualifications ?* London:Allen and Unwin.
- Pearson, I (1988) 'Tests as levers for change'. In Chamberlain, D and Baumgartner, R (eds.) *ESP in the classroom : practice and evaluation* Modern English Publications.
- Swain, M (1985) 'Large-scale communicative language testing : a case study' In Lee, YP, Fok, ACYY, Lord, R, and Low, G (eds.) *New directions in language testing* Oxford:Pergamon.
- UCLES (1992) 'Project to improve secondary English teaching - Thailand' Cambridge:University of Cambridge Local Examinations Syndicate. Mimeo.
- Wall, D and Alderson, JC (1992) 'Examining washback : the Sri Lankan impact study'. Paper presented at the Language Testing Research Colloquium, February 1992. Mimeo.
- Wiseman, S (ed.) (1961) *Examinations and English education* Manchester:Manchester University Press.

6 METHODOLOGY WASHBACK - AN INSIDER'S VIEW¹

Lam Hiu Por

Chan Sui Ki (La Salle) College

Introduction

The term "washback" or "backwash" is common in the language teaching and testing literature, i.e. tests or examinations are held to be powerful determiners of what happens in classrooms (Davies, 1968 & 1990; Forbes, 1973; Madsen, 1976; Swain, 1985; Morrow, 1986; Pearson, 1988; Hughes, 1988 & 1989; Khaniya, 1990a & 1990b; Bachman, 1990; Weir, 1990; Fullilove, 1992; Alderson & Wall, 1993; Wall & Alderson, 1993; King, 1994; see also Morris, 1972). However, it is found that not much empirical research has been conducted so far to specifically define washback, identify its nature and explain its existence (Lam, 1993: 10-38). A rare model of empirical research using direct observation as one of the tools is the Sri Lankan O-level English Language Evaluation Project (Alderson & Wall, 1992 & 1993; Wall & Alderson, 1993), which investigated the impact of the new English examination on Sri Lankan classroom teaching. The new examination was designed to reinforce the innovations introduced in textbook materials and teacher training courses. Their Impact Study consisted of a baseline study carried out at the beginning of the project and repeated two years later, a longitudinal study carried out from the beginning of 1990 to the end of 1991, and a set of questionnaires distributed over the entire four years of the study.

Apart from drawing up a series of 15 hypotheses to elaborate "washback", Alderson and Wall (1993) argue that the Sri Lankan O-level English examination has had virtually no impact on the way teachers teach although it has had a demonstrable impact on the content of lessons. They suggest the lack of impact may be due to a lack of understanding on the part of teachers of what might be an appropriate way to prepare students for the examination. They conclude that

It may even be that the exam itself - and this may indeed be true of all exams - does not and cannot *determine how* teachers teach, however much it might influence *what* they teach (1993: 127).

It could be argued that although the Sri Lankan exam was not able to determine how teachers taught in that educational setting in that particular period of time under the constraints found there. The same result may not be found elsewhere. Different educational contexts may produce different results.

This paper reports part of the results of the Use of English (UE) Research Project in Hong Kong to investigate whether examinations can determine the ways teachers teach.

The Use of English Research Project involved focussing on a revised syllabus for Use of English (RUE) which is a public examination used principally to evaluate language competence for entrance to Hong Kong's tertiary institutions. The study centred around a questionnaire administered to 61 English teachers and the analysis of a sample of textbooks. The research hypothesis at its most general level is: *Is it possible in Hong Kong to use the examination system to bring about a positive washback effect on what happens in English language classrooms? In particular, have the changes that were made to the 1989 Use of English (UE) syllabus and the way these were reflected in subsequent examination question papers made any positive contribution to the way teachers teach English and the way students learn the language in secondary six (S6) and seven (S7)?*

The research question is further expanded into nine Washback Hypotheses (WH) (compared with the fifteen WH proposed by Alderson & Wall, 1993: 120-1). They are listed below:

Timetable Washback

- (1) RUE will influence the number of periods available for its teaching within the present school **timetable**.
- (2) RUE will influence teachers' decision to allocate **a Section C period** * to prepare students for the public examination.

* Section C is the objective paper - see page 86 under Hong Kong Context

Methodology Washback

- (3) RUE will influence how teachers teach, i.e. the **methodology and methods** used to prepare students for the public examination.

Attitude Washback

(4) RUE will change teachers' attitudes.

When the social judgement theory in Social Psychology is applied (4) can be translated more specifically into:

RUE will result in the greatest amount of positive attitude change in teachers when the attitude discrepancy is intermediate (Sheri & Hovland, 1961).

(5) RUE will change students' learning attitudes.

Proofreading Washback

(6) The Proofreading question of RUE will encourage students to develop the habit of always checking their own written work for obvious errors.

Textbook Washback

(7) RUE will influence the quality of textbooks published for preparing students for the examination.

Content Washback

(8) RUE will influence what teachers teach, i.e. the content of their daily English lessons.

Performance Washback

(9) RUE will influence the actual language performance of post-1988 students in a positive direction.

Only results from question three of the teacher questionnaire are presented here to draw concluding comments on the Methodology Washback Hypothesis.

Hong Kong Context

Before 1989, the UE examination aimed at testing the ability of candidates to pursue academic courses at the University of Hong Kong (HKU). It was quite academic in nature. The examination was divided into Section I: Cursory Reading (25%, time allotted: 30 minutes), Section II: Intensive Reading (30%), Section III: Written Composition (25%) - time allotted for Section II and III was 150 minutes, and Section IV: Listening Test (time allotted: about an hour). The objective and format of the examination was radically changed in April 1989. Since then it has been used principally to evaluate language competence for entrance to Hong Kong's tertiary institutions and/or employment. The range of skills required is much wider in the revised Use of English (RUE)

examination. From 1989 to 1993, RUE consisted of four sections. An oral component for the exam has also been conducted since March 1994. The examination now has five sections: Section A: Listening (18%, time allotted: about an hour), Section B: Writing (18%, time allotted: 75 minutes), Section C: Reading & Language Systems (18%, time allotted: 90 minutes), Section D: Oral (18%, time allotted: 30 minutes - including 10 minutes for candidates to make preparation), and Section E: Practical Skills for Work & Study (28%, time allotted: 105 minutes).

The official English language admission requirement of HKU was a Grade E in the old UE exam prior to 1989. From 1989-94, HKU required what was known as a "double D" English language entry requirement, i.e. candidates had to have at least a Grade D in both Syllabus B of the HKCEE ² and the RUE examination.

RUE differs from its predecessor in many aspects (King, 1994: 17-18). In Section A: Listening, RUE is much more authentic and the tape is only played once. Candidates are required to role play tasks that resemble what people are sometimes called upon to do in their work and tertiary studies. The old listening test, however, involved playing a passage twice with the candidates unaware of the questions they would be asked. In Section B: Writing, RUE is more demanding. It requires 500 words instead of 400 in the old UE and candidates have to argue a case, defend a position or persuade an audience. Section C is the multiple-choice objective paper. The Reading part requires an in-depth understanding of an expository text rather than the former cursory reading. The Language System part aims at testing the degree the systems of the English language have been internalised by candidates. Section D: Oral, tests students' ability to give an oral presentation and take part in a small-group discussion. Section E: Practical Skills for Work & Study aims to test practical communication through simulated work and study situations and is an attempt to test reading and writing in an integrated and realistic manner.

Method and Procedure

In order to get an idea of how RUE affects the classroom teaching in sixth forms, a questionnaire was developed which covered four main areas: changes in the time allocation for the subject, the use of textbooks, teaching methodology and methods used to prepare students for the public examination, and changes in the attitudes of both teachers and students.

A preliminary version of the questionnaire was first generated in each of the four areas. It was tried out with three experienced teachers who had taught both the revised and old syllabuses. Modifications were then made to include Likert-type scales as a more objective evaluation technique to decide whether or not significant change has taken place with the introduction of the revised syllabus. Moreover, questions about the teaching of UE Oral Exam for 1994 (teaching for the Oral Exam began in 1992) were also included in the final version of the questionnaire.

Teachers were asked to complete the questionnaire in March 1993 which was made up of eight questions and a section seeking details of their teaching experience, information about their schools and their educational qualifications.

Question three of the questionnaire (see Appendix) was designed to give insights into how teachers actually teach the revised syllabus in sixth form classrooms. A variety of methods which could be used to teach each section of the UE exam were presented to teachers and they were asked to indicate the importance of these methods to them as teachers. They were asked to evaluate the methods using a five-point scale: extremely important (1), important (2), not of major importance (3), of little importance (4), and of no importance (5). Under each section, teachers were given space for individual comments.

Of the 61 teachers who completed the questionnaire, 28 of them were only familiar with teaching the revised syllabus (classified as T^{New}), while 33 teachers had taught the old syllabus and were teaching the revised syllabus (classified as T^{Old}). 34 of these teachers (56%) were taking an ILE (Institute of Language in Education) teacher training course. Thus, they represented a more progressive group in the population. 80% of the teachers had local educational qualifications while 20% of them possessed overseas qualifications (U.S., U.K., Canadian, Singaporean and Australian).

The sampled teachers worked in 42 different schools which represented 15% of the total number of schools which participated in the 1992 UE exam. In terms of exam results, the sampled schools' Grade D+ mean (i.e. Grade D³ or better) in 1992 was 53.92% which was almost the same as the population mean (53.50%).

Washback Hypothesis

The methods presented to teachers in question three of the questionnaire can be categorised into three main types: a) working through textbooks and previous exam questions; b) using non-textbook materials based on media

sources; and c) adopting an integrated approach to teaching different language skills. As these methods can be good representations of actual classroom realities (see Richards et al., 1992: 90), the Methodology Washback Hypothesis is: *RUE will influence how teachers teach, i.e. the methodology and methods they use to prepare students for the public examination*

Predicted Effects

The UE examination is not textbook-based or content-based. It is, however, a proficiency test aimed at assessing the sophistication of skills candidates can demonstrate. Specific sets of textbooks have never been recommended or prescribed by the official bodies, such as the Education Department, the Hong Kong Examinations Authority and the Curriculum Development Institute. Some pedagogical strategies have been suggested in the Use of English teaching syllabus. An overall picture of what and how teachers are expected to teach can also be inferred from the official statement about the goals of the new examination.

The aim of RUE is to foster the development of students' English language skills in order to equip them for tertiary education and/or employment. The working party which produced the new examination wanted material used in the classroom to be as authentic as possible, so that students following a UE course could see the relevance of what they were studying regardless of whether they would go on to tertiary studies or to employment. Teachers should, therefore, be aiming to bring about maximum progress in the acquisition of English and taking a longer term view than just preparing for the exam. Excessive concentration on exams is thought to be counterproductive since it is not the most efficient use of limited classroom time.

The Hong Kong Examinations Authority (HKEA) also points out that:

Use of English teachers need to put the main emphasis on improving written skills in the classroom situation, while students, apart from doing regular written work for homework, should concentrate on increasing their exposure to English through a programme of individual reading and listening at a level appropriate to their own English language standard. (HKEA, 1987: 13)

Thus the overall expectation is: teachers should maximise the classroom time for helping students with the language areas where they most need the guidance of teachers, such as Section E: Practical Skills for Work & Study. Excessive drilling of exam formats should be avoided. This is the main rationale strongly advocated by the HKEA and CDI (Curriculum Development Institute). They think that schools should not allocate a period either for the teaching of Section C (the objective paper) or the teaching of Section D: Oral. They claim the best way to teach oral skills is to incorporate an oral approach into the teaching of the whole subject (e.g. by having discussions, debates or brainstorming sessions on a topic prior to doing a composition on it). As with Section C, the HKEA and the CDI agree that students need practice to become familiar with the specialised exam format - but they claim that teachers generally exaggerate the importance of this (HKEA, 1993).

To conclude, the desirable effects which can be anticipated for teaching RUE are: a) the maximal use of authentic materials in the language classrooms; b) focusing on students' language acquisition and not excessive exam drilling; and c) adopting an integrated approach to help students develop the ability to handle realistic tasks.

Results from Question Three of the Teacher Questionnaire

Since the main focus was to find out how many teachers chose the "extremely important (1)" and "important (2)" options (referred to as Scales 1 & 2 in the data tables), these two sets of figures were added up and expressed as percentages for each of the five different sections: Section A - Listening, Section B - Writing, Section C - Reading & Language Systems, Section D - Oral, Section E - Practical Skills for Work & Study. The Likert scale means, however, were calculated by the weighted average of all five scales 1, 2, 3, 4 and 5 as reported by teachers in each group.⁴ Results are summarised in tables and inserted under relevant sections in **Discussion**.

Discussion

The discussion of the results is grouped under the teaching methods presented to teachers in question three of the questionnaire.

Method I: Working through the textbook aiming to cover most of the material it contains

	T ^{New} n=28	T ^{Old} n=33	Total n=61	
Section				
A Scales 1 & 2 (%)	32	67	51	(see Note 3)
Likert scale means	2.8	2.1	2.4	
B Scales 1 & 2 (%)	18	18	18	
Likert scale means	3.3	3.6	3.4	
C Scales 1 & 2 (%)	43	52	48	
Likert scale means	2.8	2.6	2.7	
D Scales 1 & 2 (%)	11	27	20	
Likert scale means	3.4	3.2	3.3	
E Scales 1 & 2 (%)	50	48	49	
Likert scale means	2.7	2.7	2.7	

* Scales 1 & 2 refer to those teachers who chose "extremely important" and "important", i.e. points (1) and (2) on the Likert scales.

Teachers who chose "extremely important (1)" or "important (2)" in evaluating **Method I** can conveniently be labelled as "textbook slaves", i.e. they are heavily reliant on textbooks in their classroom methodology. The findings in this section are consistent with those in question two (Textbook Washback). Teachers generally rely a lot more on textbooks in teaching Section A: Listening (51%), Section C: Reading & Language Systems (48%), and Section E: Practical Skills for Work & Study (49%) than in teaching Section B: Writing (18%) and Section D: Oral (20%).

T^{Old} teachers (i.e. with experience in teaching both syllabuses), in particular, rely more on textbooks than T^{New} teachers (i.e. only familiar with RUE) do in the listening section (67% versus 32%). The pattern is also apparent from the overall Likert scale mean in Section A (2.1 versus 2.8 - the smaller the figure, the more the perceived importance). Even if we assume that the responses are 100% honest, this textbook-reliant approach/methodology in Hong Kong is still very significant, i.e. about 50% of the teachers appear to be "textbook slaves" in teaching Section A, C and E. This is a sign of negative washback because instead of introducing more authentic materials, they prefer to use commercial textbooks, most of which are basically modified copies of the exam papers (see Fullilove, 1992). This textbook-reliant situation may be assumed to be worse than the data indicate because 56% of teachers filling in the questionnaire were on ILE courses and can be taken to be more progressive than the average teacher. When filling in questionnaires, there is also a subconscious tendency to answer like a "model teacher" - teachers are shrewd and know how they

ought to answer. It is debatable whether the fact that the respondents knew the questionnaire would be anonymous led to complete honesty.

Only about 20% of teachers conform to the definition of "textbooks slaves" in the writing (B) and oral (D) sections. It is interesting to discover that T^{Old} view textbooks in the oral section as more important than their younger, less experienced counterparts do (27% versus 11%).

Method II: Working through questions from previous examinations or setting your own questions that parallel those found in recent examinations

Section	T^{new} n=28	T^{Old} n=33	Total n=61
A Scales 1 & 2 (%)	68	82	75
Likert scale means	2.1	2.1	2.1
B Scales 1 & 2 (%)	54	45	49
Likert scale means	2.5	2.7	2.6
C Scales 1 & 2 (%)	61	55	57
Likert scale means	2.3	2.6	2.5
D Scales 1 & 2 (%)	29	67	33
Likert scale means	2.7	2.7	2.7
E Scales 1 & 2 (%)	79	76	77
Likert scale means	1.9	2.1	2.0

In **Method II**, the "exam slave" situation turns out to be more serious than the "textbook slave" phenomenon. Teachers view practising examination questions as very important in Section A (75%) and Section E (77%); important in Section C (57%) and even in Section B (49%) and Section D (33%). The differences between T^{Old} and T^{New} are again very revealing. For example, in their responses in the listening section (82% versus 68%) and oral section (67% versus 29%), T^{Old} are found to be much more examination-oriented. They believe the best way to prepare students for exams is by doing past questions. This can be regarded as the classic "bad washback" of exams on teaching (see Forbes, 1973 & Madsen, 1976), since teachers are expected to bring about maximum progress in students' acquisition of English and to take a longer term view than just preparing for the exam. In addition, excessive concentration on exams is held to be counterproductive as it does not make the most efficient use of limited classroom time.

Method III: Using non-textbook material based on media sources

	T^{New} n=28	T^{Old} n=33	Total n=61
Section			
A Scales 1 & 2 (%)	71	58	66
Likert scale means	2.3	2.6	2.5
B Scales 1 & 2 (%)	93	82	84
Likert scale means	1.7	1.9	1.8
E Scales 1 & 2 (%)	46	55	51
Likert scale means	2.6	2.5	2.5

The most important use of **Method III** is in Section B (84%), followed by Section A (66%) and Section E (51%). T^{Old} , however, are less progressive than T^{New} in this regard. They view this method as less important than T^{New} do in both the listening section (58% versus 71% - Likert scale means: 2.6 versus 2.3) and the writing section (82% versus 93% - Likert scale means: 1.9 versus 1.7).

Method IV: Producing learning activities resulting from class debates, discussions, brainstorming sessions

	T^{New} n=28	T^{Old} n=33	Total n=61
Section			
B Scales 1 & 2 (%)	82	79	80
Likert scale means	2.0	1.9	2.0
D Scales 1 & 2 (%)	61	45	51
Likert scale means	2.2	2.4	2.3

Teachers view **Method IV** as more important in Section B (80%) than in Section D (51%). T^{Old} , are less progressive than T^{New} since they perceive this method as less significant in both the writing section (79% versus 82%) and in the oral section (45% versus 61%).

Method V: Adopting an integrated approach in teaching Listening & Writing

	T ^{New} n=28	T ^{Old} n=33	Total n=61
Section			
A Scales 1 & 2 (%)	61	33	44
Likert scale means	2.5	3.0	2.8
B Scales 1 & 2 (%)	79	39	41
Likert scale means	2.8	2.7	2.7

Teachers perceive **Method V** as important in Section A (44%) and Section B (41%) but their importance is less when compared with the methods in III and IV. T^{Old} and T^{New} differ in their views greatly in using this method in Section A (33% versus 61%) and in Section B (39% versus 79%).

Method VI: Ignoring the Section C and Section D apart from the minimum practice required to ensure familiarity with the exam question formats

	T ^{New} n=28	T ^{Old} n=33	Total n=61
Section			
C Scales 1 & 2 (%)	32	15	23
Likert scale means	3.0	3.4	3.2
D Scales 1 & 2 (%)	25	6	15
Likert scale means	3.0	3.8	3.4

Method VII: Giving practice in report/memo/formal-letter/minute writing

	T ^{New} n=28	T ^{Old} n=33	Total n=61
Section			
E Scales 1 & 2 (%)	79	94	87
Likert scale means	1.9	1.8	1.9

In **Method VI** only a few teachers dare not to practise Section C (23%) and Section D (15%). These results are consistent with the findings in question 5(a) (Timetable Washback) in which 69% of teachers reported they allocated a Section C period in either S6, or S7, or in both S6 and S7 (see Lam, 1993). The Washback Hypothesis which states that tests that have important consequences will have washback effects is verified (see Alderson & Wall, 1993: 120). In preparing students for coping with Section C and D of the exam, the hypothesis held by most teachers can be expressed as: *these two sections bear high scores (18% each) and therefore we must give more practice sessions to students*. In their view, not doing that may mean "bad" results - as expressed by a T^{Old} teacher in question 5(a).

The negative washback is much more serious for T^{Old} than T^{New} in Section C (15% versus 32%) and in Section D (6% versus 25%). Results in question 5(a) again confirm this distinction, i.e. T^{Old} tend to be more likely to allocate a Section C period both in S6 and S7 than T^{New} (70% versus 50%). This also explains why T^{Old} reported giving more practice to students in **Method VII: report/memo/formal-letter writing** than T^{New} (94% versus 79%) in Section E - this can be taken to be another sign of negative washback.

Other strategies described by individual teachers

Percentage of teachers who gave response:

Section

A	18
B	20
C	23
D	16
E	7

Some teachers' comments under **other strategies (please described)** in question three are revealing. Some of the progressive teachers added their own ideas (18%, 20%, 23%, 16% and 7% in Section A, B, C, D and E respectively) - they found the given options too limiting and so added descriptions of their own creative methods:

"ask students to hold interviews with native-speakers and tape the interviews" (quoted from a T^{New} teacher under Section A: Listening)

"work with those doing B.A. TESOL in Hong Kong City Polytechnic to organise activities that motivate students to talk" (quoted from a T^{New} teacher under Section D: Oral)

"writing book report/group project/analysing the language skills of good essays by prominent writers" (quoted from a T^{Old} teacher under Section B: Writing)

"student attend extra oral lessons conducted by native speakers" (quoted from a T^{Old} teacher under Section D: Oral)

Creative methods are evidence of good exam washback - they indicate that not all teachers are forced into the traditional stereotype of Hong Kong teachers whose methods are totally exam-dominated. Thus, RUE has good washback effects for some teachers, but not for others [compared with WH(15) in Alderson & Wall, 1993: 121].

Conclusion

The Methodology Washback Hypothesis can be analysed in several dimensions. Unlike the situation in Sri Lanka, there is evidence of both positive and negative washback in Hong Kong. Signs of negative washback include a heavy reliance on textbooks and past exam question papers, as well as the allocation of isolated teaching periods for teaching the objective test paper (Section C) and the oral paper (Section D). Positive washback is evidenced by teachers creating more authentic materials from the mass media, producing meaningful learning activities, encouraging more student participation, and adopting an integrated approach when teaching listening and writing. It also includes other innovative methods reported by teachers.

The findings also reveal some differences between the experienced and the less experienced teachers. The more experienced teacher group are consistently more likely to be subjected to negative washback and less likely to exhibit positive washback than their younger, less experienced counterparts. The differences suggest that they do not share the same teaching attitudes and philosophy, and may represent two slightly different cultures among the UE teachers in Hong Kong. These findings may assume that the more experienced UE teachers are less likely to be affected by syllabus innovations because they are more set in their ways and more confident of themselves as a result of more years of experience and the fact that they are more realistic in assessing what is functional in their working situations, i.e. they are more likely to assess their performance on the results of their students in the public examinations rather

than on the quality of their teaching (see Richards et al., 1992: 83; Morris, 1990). It seems it isn't enough just to change/structure the exam in a certain way. The challenge is to change the teaching culture: to open teachers' eyes to the possibilities of exploiting the exam to achieve positive and worthwhile educational goals.

To sum up, the changes that were made to the 1989 Use of English syllabus and the way these have been reflected in subsequent examination papers have made some positive contribution to the way teachers teach English in the classroom. Unfortunately this contribution is severely limited because of the exam-oriented English teacher culture which is influenced by many complex issues, such as teachers' language competence and personalities, the changing student learning culture, as well as school and educational politics.

Notes

- 1 The writer was educated in Hong Kong and at the University of Leeds. He experienced the old Use of English examination as a student and has taught the new Use of English syllabus in different secondary schools in the territory.
- 2 The Syllabus B version of the English language paper of the HKCEE(Hong Kong Certificate of Education Examination) is taken by Secondary five students who study in Anglo-Chinese schools in which English is the official medium of instruction. The vast majority of secondary five students in Hong Kong opt for the Syllabus B examination and not the easier alternative, Syllabus A, which is aimed at students studying in Cantonese-medium secondary schools. A Grade C or above in the HKCEE(Syllabus B) is equivalent to an O-level pass.
- 3 Prior to 1995, Grade D(08) has been the critical result in the UE examination because HKU and the University of Science & Technology have required this for entry purposes. The existing grading system for UE passing grades are A(01), A(02), B(03), B(04), C(05), C(06), D(07), D(08), E(09) and E(10), while F(11), F(12) and Unclassified are the failing grades. Hence, Grade D(08) has been the lowest acceptable grade for admission to the two universities referred to. Prior to 1989, when the old UE exam was in place, the official English language admission requirement of HKU was a Grade E.
- 4 For example, the calculations for the percentages of "Scales 1 & 2" under T^{New} , T^{old} , and Total in the annex under *Method I: Working through the textbook aiming to cover most of the material it contains* in Section A: Listening are as follows:

$$T^{\text{New}}: \text{No. of teachers who chose scales 1 and 2} / 28 \times 100\% \\ = (2 + 7) / 28 \times 100\% = 32\%$$

$$T^{\text{old}}: \text{No. of teachers who chose scales 1 and 2} / 33 \times 100\% \\ = (7 + 15) / 33 \times 100\% = 67\%$$

$$\text{Total: Total no. of teachers who chose scales 1 and 2} / 61 \times 100\% \\ = (9 + 22) / 61 \times 100\% = 51\%$$

The Likert scale means of Method I in Section A are calculated as follows

$$T^{\text{New}} : \text{No. of teachers who chose scales 1, 2, 3, 4 \& 5 / 28} \\ = (2 \times 1 + 7 \times 2 + 16 \times 3 + 2 \times 4 + 1 \times 5) / 28 = 2.8$$

$$T^{\text{old}} : \text{No. of teachers who chose scales 1, 2, 3, 4 \& 5 / 33} \\ = (7 \times 1 + 15 \times 2 + 9 \times 3 + 2 \times 4 + 0 \times 5) / 33 = 2.1$$

$$\text{Total: Total no. of teachers who chose scales 1, 2, 3, 4 \& 5 / 61} \\ = (9 \times 1 + 22 \times 2 + 25 \times 3 + 4 \times 4 + 1 \times 5) = 2.4$$

References

- Alderson, J. C. & Wall, D. (1992) *The Sri Lankan O-level evaluation project: Fourth and final report*. Lancaster University.
- Alderson, J. C. & Wall, D. (1993) Does washback exist? *Applied Linguistics*. Vol. 14, No. 2: 115-129. Oxford University Press.
- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford University Press.
- Boyle, J. & Falvey, P. (eds) (1994). *English language testing in Hong Kong*. Hong Kong: Chinese University Press.
- Chamberlain, D. & Baumgardner, R. (eds.) (1988). ESP in the classroom: Practice and evaluation. *ELT Document 128*. Modern English Publications.
- Davies, A. (1968). *Language testing symposium: A psycholinguistic approach*. Oxford University Press.
- Davies, A. (1990). *Principles of language testing*. Blackwell.
- Forbes, D. (1973). Selling English short. *English Language Teaching Journal*. 27: 132-137.
- Fullilove, J. (1992). The tail that wags. *Institute of Language in Education* Vol. 7: 131-147.
- Hong Kong Examinations Authority (1988). *New Use of English*.
- Hong Kong Examinations Authority (1993). Consultation with the HKEA HKEA Sources.
- Hughes, A. (1988). Introducing a needs based test of English language proficiency into an English medium university in Turkey. In Hughes (ed.)

- Hughes, A. (ed.) (1988). Testing English for university study. *ELT Document 127*. Modern English Publications.
- Khaniya, T. R. (1990a). Examination as instruments for educational change: Investigating the washback effect of the Nepalese English exams. PhD thesis, University of Edinburgh.
- Khaniya, T. R. (1990b). The washback effect of a textbook-based test. *Edinburgh Working Papers in Applied Linguistics* 1: 48-58.
- King, R. F. (1994). Historical survey of English language testing in Hong Kong. In Boyle, J. & Falvey, P. (eds), pp. 3-29.
- Lam, H. P. (1993). Washback - Can it be quantified? A study on the impact of English examinations in Hong Kong. MA thesis, University of Leeds.
- Lee, Y. P., Fok, C. Y. Y., Lord, R. & Low, G. (eds.) (1985) *New direction in language testing*. Pergamon Press.
- Madsen, H. (1976). New alternatives in EFL Exams or, How to avoid selling English short. *English Language Teaching Journal* 30: 135-144.
- Morris, B. (1972). *Objectives and perspectives in education: Studies in educational theories*. Routledge and Kegan Paul.
- Morris, P. (1990). *Curriculum development in Hong Kong*. The Faculty of Education. University of Hong Kong.
- Morrow, K. (1986). The evaluation of tests of communicative performance. In Portal (ed.).
- Pearson, I. (1988). Test as levers for change. In Chamberlain & Baumgarden (eds.): 98-107
- Portal, M. (1986). *Innovations in language testing*. NFER/Nelson.
- Richards, J. C., Tung, P. & Ng, P. (1992). The culture of English language teacher: a Hong Kong example. *RELC Journal*. Vol. 23, No. 1: 81-102.
- Sherif, M. & Hovland, C. (1961). *Social Judgement*. Yale University Press.
- Swain, M. (1985). Large-scale communicative testing. In Lee et al. (eds.).

- Wall, D. & Alderson, J. C. (1993) Examining washback: the Sri Lankan impact study. *Language Testing*. Vol. 10, No. 1: 41-69.
- Weir, C. (1990). *Communicative language testing*. Prentice Hall International English Language Teaching.

Appendix

QUESTION THREE IN THE USE OF ENGLISH QUESTIONNAIRE

Please indicate the importance of the following methods to you as a teacher. Please rate their actual importance using this five-point scale:

1.....2.....3.....4.....5

extremely importance	important	not of major importance	of little importance	of no importance
-------------------------	-----------	----------------------------	-------------------------	---------------------

LISTENING

- working through the textbook aiming to cover most of the material it contains
- using non-textbook material taped from radio or television
- working through tapes from previous examinations
- not focusing directly on listening tasks but covering this aspect in integrated tasks that involve other skills such as reading and writing
- other strategies (please describe)

WRITING

- working through the textbook aiming to cover most of the material it contains
- working through questions from previous examinations or setting your own questions

that parallel those found in recent examinations

- producing your own writing tasks based on media sources
- producing your own writing tasks resulting from class debates/discussions/ brainstorming sessions
- not focusing directly on the type of writing tasks found in Section B, but covering this aspect in integrated tasks such as those needed for Section E
- other strategies (please describe)

READING & LANGUAGE SYSTEMS

- working through the textbook aiming to cover most of the material it contains
- working through questions from previous examinations or setting your own questions that parallel those found in recent examinations
- ignoring the Section C exam paper apart from the minimum practice required to ensure familiarity with the exam question formats
- other strategies (please describe)

ORAL

- working through the textbook aiming to cover most of the material it contains
- working through questions that parallel those found in the sample paper from the HKEA
- ignoring this paper in the sense that no periods are timetabled for it, but instead relying on class debates/discussions/brainstorming sessions in connection with teaching the other skills that the exam will be testing.
- ignoring this paper apart from the minimum practice required to ensure familiarity with the exam question formats
- other strategies (please describe)

PRACTICAL SKILLS FOR WORK & STUDY

- working through the textbook aiming to cover most of the material it contains
- working through questions from past exam papers or using your own questions that parallel those in recent papers
- setting writing tasks based on material taken from newspapers and magazines which aim in a general way to improve the enabling skills that this section requires
- giving practice in report/memo/formal-letter/minute writing
- other strategies (please describe)

7 TOWARDS A BROADER VIEW OF AUTHENTICITY IN THE LANGUAGE CLASSROOM: A DISCUSSION PAPER

Patricia Galien & Wendy L. Bowcher
Obirin University, Tokyo

Introduction

In an effort to provide students with learning situations which embrace the complexities of language, ESL teachers and curriculum developers have been increasingly working toward creating a communicative language classroom which provides students with exposure to authentic language as well as opportunities for genuine communication. This desire has resulted in an increase in the amount of authentic resources that teachers may take into the classroom. However, although the definition of authentic resources is clear, authenticity in terms of language learning tasks and communication in general appears to be much less clear. At times this lack of clarity results in classroom tasks which do not reflect the teacher's original desire for authentic communication. This paper is part of an ongoing evaluation of an EFL programme, the aim of which is to investigate what is meant by authentic communication, and to develop ways of creating truly authentic communication in the classroom. The focus of this paper is the first step of that evaluative process. It will firstly review some of the linguistic theories that have had significant impact on language education, and will raise issues about the nature of authenticity with specific reference to systemic functional linguistic theory. It is hoped that through this investigation the path to a broader definition of authenticity in the language classroom will be developed and the creation of a more enhanced learning situation will result.

Theories of Language and Their Impact on Language Education

All language teaching reflects some conscious or unconscious view of language, and these views or beliefs about the nature of language have been formed largely by the linguistic theories of the time. For many years language teaching methodology was greatly influenced by traditional grammars which viewed language as a system of rules. As a result, language teachers centred their lessons around structurally-graded syllabi firmly believing that if the grammar of the language were learned, communication could take place. Language was seen as a set of rules with the assumption that if students learned the rules they would know the language.

Transformational generative grammar is also based on the assumption that language is a system of rules which are known by speakers of a language and used to form, or generate grammatical sentences. Chomsky's generative grammar offers a complex linguistic system but as in traditional grammars, places emphasis on the underlying linguistic system, or the ideal, while ignoring the actual social activity of communicating. According to Chomsky, in the technical sense, linguistic theory is mentalistic, since it is concerned with discovering a mental reality underlying actual behavior. Observed use of language or hypothesized dispositions to respond, habits, and so on, may provide evidence as to the nature of this mental reality, but surely cannot constitute the actual subject matter of linguistics, if this is to be a serious discipline (1965:4).

An important influence that Chomsky's generative grammar provided for language education was the distinction between linguistic competence, (the underlying knowledge of one's language) and linguistic performance, (the actual use of language). Language teaching syllabi which accepted this dichotomy placed emphasis on learning grammar rules in an effort to develop linguistic competence. It was generally believed that successful performance resulted from the internalization of these rules and the ability to generate, or "construct novel utterances ... in conformity with the grammatical system of the new language" (Rivers 1981: 79). This approach, which extended the structuralist stance of traditional grammar, did not however take into account other factors influencing language use - factors of a social nature.

With the influence of sociolinguistic theories as well as ethnographic and anthropological research into the role of language in society (Labov 1969, 1970, 1972; Hymes 1971, 1972, 1974; Firth 1957; Halliday 1973, 1975, 1978) came the development of 'communicative' syllabi and communicative approaches to language teaching. Early examples of communicative syllabi and approaches are the notional-functional syllabus (Wilkins 1976), needs-based approaches to curriculum design such as that by Munby (1978) and the Threshold Level approach of the Council of Europe (1973). The focus had shifted from viewing language as a "box" of rules, or structurally, to viewing language in terms of how it is used in social contexts. Thus, rather than teaching a particular grammatical point such as the past tense, teachers were now teaching language functions such as, "ordering in a restaurant", or "giving directions", and notions such as "telling time". While such syllabi and approaches provided important steps in the incorporation of language as a means of communication in social contexts, when applied to the development of a language education syllabus, one type of list was replaced with another. Nunan (1988) notes that:

"Widdowson pointed out as long ago as 1979 that inventories of functions and notions do not necessarily reflect the way languages are learned any more than do inventories of grammatical points and lexical items. He also claims that dividing language into discrete units of whatever type misrepresents the nature of language as communication." (Nunan 1988:37).

An important step in language curriculum design was the notion of sociolinguistic rules, or rules of speaking which are part of what Hymes (1971) has called communicative competence. According to Hymes, there are communicative rules that are acquired as the mother tongue is learned. He states that along with a system of grammar, a child learns a system of language use which is part of his or her process of socialisation within a speech community (1971). In the field of sociolinguistics where language is viewed as a system of variation determined by certain contexts of situation, it soon became important that "second language instruction should be to facilitate learners' acquisition of communicative competence, the ability to speak both accurately and appropriately." (Wolfson 1989:36)

Important steps have been made in applying the notion of communicative competence to the language classroom. Canale and Swain (1980) and Canale (1983) have proposed a framework for communicative competence with the object of preparing and encouraging learners "to exploit in an optimal way their limited communicative competence in the second language in order to participate in actual communication situations" (1983:17). In order to fulfil these objectives, a variety of teaching methods have been developed, in particular a movement away from context-free grammar exercises towards attempts to guide students to communicate in meaningful ways (Wolfson 1989:48). Rivers (1981) suggests:

"...that communicative rules can be learned by the non-native speaker through a period of immersion ...[or] where such immersion is not possible, the teacher must be aware of the needs to develop this ability to communicate acceptably according to the sociolinguistic rules of the speech community, and must include this 'organisation of diversity' as an essential element of the program." (1981:237)

In the mid-sixties M.A.K Halliday called for a deeper appreciation of aspects of language which he referred to as "extra-linguistic". He states that:

"In investigating grammar and phonology, linguists have tended to insist on the autonomy of their subject; this is natural and useful, since these are the 'inner' strata of the linguistic system, the core of language so to speak... But they are in turn contingent on other systems which do relate outside language. Moreover we take the view that we can understand the nature of the inner stratal systems of language only if we do attempt to relate language to extra-linguistic phenomena." (revised version of the original in Brumfit and Johnson, 1979: 44)

Unlike traditional and generative grammarians, Halliday, when looking for explanations of linguistic phenomena, explains linguistic phenomena in terms of the social system. He states that language is "the main channel through which the patterns of living are transmitted" to a child _ it is through language that a child "learns how to act as a member of a 'society' and to adopt its 'culture', its modes of thought and actions, its beliefs and its values" (1978: 9). Halliday rejects as unnecessary the competence-performance dichotomy of Chomsky (1965) and the communicative competence-performance dichotomy of Hymes (1971/79). Instead, Halliday views the linguistic system as a potential from which linguistic choices are made according to the context of situation in which the speaker, or writer, finds him/herself. He states that "there is no need to bring in the question of what the speaker knows; the background to what he does is what he could do - a potential, which is objective, not a competence, which is subjective" (1978:38).

Although communicative language teaching is widely accepted by the majority of language teachers, the practical methodology of incorporating the complexities of language is still being developed. Literature concerned with ESL indicates some confusion as to how to implement an approach which takes into account so much of the complexity of language. Much research in the field of communicative language teaching has indicated that many of the tasks that are set for students are in practice not authentically communicative. Anne Burns (1990) refers to research by applied linguists which has demonstrated that:

"...many so-called communicative classrooms, despite incorporating tasks which now form part of a 'communicative' teaching repertoire - pair work, group work, jigsaw or information gap activities - are in fact based on traditional patterns of classroom interaction and amount to little more than drills, grammatical explanations and a focus on accurate forms." (1990: 37).

Crookall (1990) has also found that in many schools some more traditional methods are still employed, and observations of actual classroom behaviour (Seliger & Long 1983) reveal that communicative language teaching is sometimes "more an ideal than a reality, even for teachers who are aware of communication as a real goal" (1990: 12). Crookall also points out that even when teachers want to use communicative activities they are often confused as to how to implement them (1990: 12). As a result, a teacher may desire to set up a communicative task which is centred on a structural form or on a function, but in actuality end up merely giving it a communicative cloak. An important step in developing authentically communicative tasks in the language classroom then, is to have a clear understanding of language and its role in communication. In other words, an understanding of language as social behaviour.

A University English Language Programme (ELP) and an Incorporation of Authenticity

In the English Language Programme which is the subject of this paper, emphasis is placed on communication in the development of academic skills, as well as in the development of more general language ability. The students enter the programme with about six years of English language study at the high school level. The high school English course has provided them with a fair amount of structural practice, but very little experience at putting English to communicative use. Their level is generally regarded as being intermediate.

The ELP is a two year intensive English programme of required language study for all students who enter the university. During their first year, students have six 90 minute classes in the ELP, while during their second year they have four. The English language programme has evolved into a content-based programme, the content of which is theme-based with integrated skills. This integration is achieved by teaching-teams of three teachers for each class. The team divides the integrated materials not by skill, but in accordance with their assessment of the class. An example of a unit is the Education Unit. The content and skills for this unit of work are summarised in table 1.

Table 1. Sample Unit of Work from the Semester Schedule for First Year ELP

Week	Theme	Overall Goals	Reading/ Vocabulary	Writing	Speaking/ Functions	Listening	Grammar
9 June 5-9	Education	<u>Compare/ Contrast</u>	Rdg & Note-taking Main Idea	Comparison and contrast essay Thesis Statement Introductory paragraphs	Requesting (repetition, clarification, info) Group Discussion	Details/ Note-taking Homework gist, details, notetaking	Superlatives Comparison "More" Expression

Presently the programme makes use of authentic material extensively, but not exclusively. For example, in the Education Unit the students read a newspaper article which compares education in Japan with education in Great Britain. They also watch a video documentary on university entrance in the U.S. and then, through group discussion, compare and contrast the contents of this video with their own experience of university entrance in Japan.

The ELP is continually being evaluated by its members and there is a desire for increased authenticity. Discussions amongst staff about authenticity, however, have highlighted two areas connected to communicative language teaching which it appears will need to be further clarified and investigated in order to continue to utilise authentic materials. The first is to have a clear understanding of the nature of language and its role in society, and the second is the need to evaluate what is done in the light of this understanding - to recognise where the teaching methodology is moving towards communication and where it is not, and what this means for the second language classroom in the context of the ELP.

Most teachers agree that the aim of the programme is to provide students with as many opportunities to interact in a variety of social settings, and rather than on the underlying linguistic rules of grammar, focus should be placed on language as a means of communicating in social contexts. It is felt that the linguistic theory which offers the most useful framework of language as a social phenomenon is the systemic functional linguistic theory, and that a closer study of the nature of language within this theoretical framework will better enable an investigation of authenticity.

Defining Authenticity with Reference to Systemic Functional Linguistics

In systemic functional linguistic theory, language is viewed as social activity. In this theory, language is related "primarily to one particular aspect of human experience, namely that of social structure" (Halliday & Hasan 1985: 4). Language is used to do many things in people's lives and it is intimately related to their lives. It is used to direct, to instruct, to hurt, to love, to fear, to destroy. Furthermore, language is a system of choices driven by the social contexts in which it takes place.

Within the systemic functional linguistic model, language is viewed in relation to its context of situation. It is important at this stage of the discussion to understand the notion of context and the part that it plays in the accompanying language. What underlies the coining of the term *context* is the understanding of language as text. In systemic functional linguistics, text refers to any instance of spoken or written language. The term context, derives from two words, *con* and *text*, so *context* means *that which accompanies text*. In a sense, text and context are two ends of a scale. That is, an understanding of the context, means that one can predict to a certain degree the language that will be used. Conversely, understanding the language that is taking place, gives an insight into the context of situation. The relationship of the text to its context is functional rather than one of physical proximity. This points to two different types of situation. There is what Hasan (1980) refers to as the material situational setting and there is the context of situation. The material situational setting may or may not include elements that are part of the context of situation. In some instances, it may be largely irrelevant. An example of such an instance may be two students gossiping to each other in the university library. The material situational setting in which this takes place, may have no bearing on the talk which takes place. There will be other features of the context of situation, however, which will have some bearing. These may include the relationship between the two students, the people they are gossiping about etc. There are of course situations in which the material situational setting places language in a subsidiary role. An example of such an instance may be where the same two students are asking the librarian about the location of certain books. The librarian may be referring to a catalogue, or the floor on which the books may be located, or whether or not the books have already been borrowed. In the first example, features of the material situational setting are unlikely to be "picked up as the ingredients of the context of situation" (Hasan 1980: 110), whereas in the second example they most probably will be. The part that context of situation plays in determining what choices are made in the linguistic system is the focus of the following section, as it is this aspect of language which is believed to have been largely overlooked in the ELP programme, and which has an important bearing on its treatment of the concept of authenticity.

In systemic functional linguistic theory, the context of situation has three components: field, tenor and mode. Field refers to what is happening in the context of situation, or the subject matter in which the participants are involved, and to the nature of the social action that is taking place. Tenor refers to the kinds of relationships of the individuals taking part in the text. Mode refers to the role that language is playing in the situation, which includes what the participants are expecting the language to do for them in that situation. It also includes the "symbolic organisation of the text, the status that it has, and its function in the context, including the channel (is it spoken or written or some combination of the two?) and also the rhetorical mode, what is being achieved by the text in terms of such categories as persuasive, expository, didactic, and the like" (Halliday & Hasan 1985: 12).

An important aspect of Halliday's interpretation of the context of situation is that field, tenor and mode are not kinds of language use or components of the setting in which language occurs, rather, together, they make up the "conceptual framework for representing the social context as the semiotic environment in which people exchange meanings" (1978: 110). Meanings are exchanged through language within a context of situation. Without the context of situation a great deal of the meaning, if not all of it is lost. In order to define authenticity, therefore, utilising the conceptual framework of context of situation is crucial.

In drawing on Halliday's description of the components of the context of situation, it was felt that there were two major dimensions of authenticity. They are, authenticity of materials, and authenticity of communication within a context of situation. Within the context of situation there are three dimensions: the participants, the subject matter and the activity in which the participants are engaged; the purpose; and the role of the participants in the situation. These dimensions are similar to Halliday's original description of the context of situation, but have been simplified in order to arrive at an understanding of authenticity which is applicable to the language classroom.

Authentic Materials

- Authentic materials are defined as those which are not designed with the purpose of language teaching but are intended for native speakers (Porter et al. 1987: 177). They have been created for a genuine communicative purpose and occur within their appropriate contexts of situation. Authentic materials often used in the classroom include such resources as "newspaper articles, magazine articles, television features, stories, cartoons, real signs and notices of public announcements, menus, and the like" (Scarino et al. 1988: 22). These authentic

resources represent language varieties with distinct schematic structures, and patterns of linguistic choices that have been made and which are appropriate to the contexts in which they occur. These materials provide great potential for the second language learner in that they are an exposure to realia - an exposure to language outside the language classroom. A teacher's understanding of the contexts, structures and linguistic features of these authentic materials can be an extremely useful tool in teaching language and in exposing students to varieties of English. Nunan (1994) further suggests that non-authentic materials may act against the long term goals of learners by misrepresenting the language, since non-authentic materials may be models of language which do not reflect language heard outside the classroom.

In their discussions of authentic materials, Harmer (1983) and Scarino et al. (1988) have stated that when people use informational resources outside the classroom they do so because they have a purpose. Either the material is intrinsically interesting and they are listening or reading in order to collect information, or they are using the resource with some extrinsic purpose in mind. The same is true of the language learner. Within the context of the language classroom however, the resource may have relevancy of purpose simply because the student finds it interesting, or the resource may help the student to achieve some extrinsic goal

Understanding the needs of the students is helpful in determining what resources will be more likely to be used in the contexts in which the students will be later encountering, but this is not enough in developing an authentically communicative classroom. A resource which is taken out of its original context, which is then not given relevance in the context of the language classroom or which is not used for an authentically communicative purpose, is a resource which has lost much of its potential for enhancing language learning.

Context of Situation - The Educational Context

In relating language to the educational context, Halliday states that "learning is, above all, a social process [and that] knowledge is transmitted in social contexts" (1985: 5). If language is a social activity and learning is also a social activity, then the classroom itself provides an authentic situation through which authentic communication can be developed.

As an authentic situation, the classroom embodies varieties of English that it is important for the students to become familiar with if they are to succeed in taking part in the activities of the classroom. One example of an "important source of communicative data for classroom learners is 'teacher talk' in the

target language" (Scarino et al. 1988: 23). This communication may take the form of socialisation, classroom management, negotiating activities, feedback, topic discussion, explanation of language phenomena, story telling, anecdotal comments and the issuing of commands. Many teachers and researchers agree that "conversation and instructional exchanges between teachers and students - when authentic and relevant - can provide the best opportunities for learners to practice L2 skills, to test out their hypotheses about how the language works, and to get useful feedback" (Ernst 1994: 294).

Teacher talk is one example of authentic communication which has been widely studied, but the language classroom also offers students many other opportunities to talk with one another in genuinely communicative ways. Students socialize with one another, share their ideas and work. They also give feedback to each other on speaking or writing activities. They may help each other understand new vocabulary, rhetorical patterns, ideas and information. All of these communicative tasks are ones which are natural to the classroom context and comprise authentic communication. However, while for native speakers the language of the classroom is readily understood, for the second language learner there may need to be some explicit instruction in this authentic use of language. An important point here is that an awareness of the possibilities of the classroom as an instance of authentic communication can be rewarding to both the teacher and to the students. In the communicative classroom many methods are used with the purpose of preparing students for communication in the target language. While some of these methods are classroom oriented and do not naturally occur outside the classroom, this does not mean that they are not genuinely communicative. Those tasks which are given a clear context, purpose and made relevant to the student through an awareness of the communicative roles of participants, will more likely be effective authentic language learning tasks.

The examples of communication which take place in the situation of the language classroom listed above are clearly appropriate and authentic. The question of authenticity becomes less clear when one examines the language goals of activities which are commonly connected with communicative language teaching. One of the major aims of the communicative classroom is to prepare students to go out of the classroom and to communicate in authentic contexts in the target language. The next section examines the notion of authenticity of communication in general in relation to the kinds of information that teachers need in order to make their language teaching more authentic.

Context of Situation - Communication in General

In defining authenticity, it is imperative to note that where genuine, or authentic communication (or indeed miscommunication) takes place there is always a context of situation. The context of situation determines the linguistic choices that are made in relation to the subject matter, the activities taking place, the statuses or roles of the participants in the situation, the channel of communication, and the overall purpose of the situation itself, i.e. those linguistic choices which move the situation along towards a certain goal. The linguistic choices that are made derive from the alternatives available in the language. In the language classroom, if decontextualised language is taught, or if students are not provided with the tools by which they can learn to understand that there are choices which are contextually appropriate, then students are being denied the richness of the second language, and indeed, the culture of which this language is a part. Therefore, as language is never void of context, the ELP programme's definition of authenticity must include some notion of context.

An overriding need of the ELP students is to behave linguistically appropriately in the target language, and to most effectively put language to use in communicating in the many and varied authentic contexts which they will encounter. Not only do students need to be exposed to a range of contexts of situations, they need to know how the language fulfils the goals of the social situation. This does not mean stripping language down to some basic structures, but means providing students with access to a range of linguistic choices which can be made in order to reach those goals. The notion of authenticity must therefore include the notion of context of situation.

Participants, Subject Matter and Activity

There is a two-way relationship between the context of situation and the linguistic elements which reflect the purpose, participants, subject matter and activities in which the participants are engaged. One does not necessarily precede the other. Halliday states that "we do not, in fact, first decide what we want to say, independently of the setting, and then dress it up in a garb that is appropriate to it in the context... The 'content' is part of the total planning that takes place" (1978: 33). Just as in the classroom one can predict to a reasonable degree the kinds of activities which will take place and the subjects that will most likely be covered, in the wider social sphere there are certain activities which more commonly take place in certain contexts of situation and certain subject fields that these activities accompany. The material setting may influence to a large degree the subject matter of the talk that will take place, or

the field of activity, such as science, or sport, may influence the subject matter of the talk irrespective of the setting. An understanding of the relationship between the context of situation, the subject matter, the activities which take place and the participants who engage in these activities is an important consideration in defining authenticity in the language classroom.

Purpose

The uses to which we put language are determined by the social activities in which we engage. For native speakers of the language, the relationship between context of situation and the purpose or goal of the situation, and hence of the language used in the situation, is often held at an unconscious level. Native speakers of a language do not usually say, "Today I'm going to use language to buy some train tickets", or "to annoy my husband", or "to tell a joke", or "to show that I understand the documentary that I have just watched". These are things that are done through language at an unconscious level. There are times, however, when the purpose for which language is used does become explicit. An example of this would be a conflict in the office when someone has been particularly rude to a co-worker. The co-worker may say, "I'm going to tell him exactly what I think!" In this situation the purpose of language has been brought to a conscious level. Another example may be where someone has asked a friend to talk with another friend in order to console that other friend. Again, the purpose of the language has been made explicit. There are often conscious goals for language and contexts of situation are occasionally consciously set up. However, while the native speaker knows how to control language at both the conscious and the unconscious level, the second language learner often has to struggle with his/her limited understanding of the language and the ways in which it varies from situation to situation. In the language classroom, therefore, the purpose for which language is being used needs to be explicitly taught, and consciously developed (Carter 1994). This may entail the setting up of a variety of contexts of situation and the explicit teaching of the underlying purpose of these contexts of situation along with the appropriate linguistic tools which native speakers use in fulfilling these purposes. What is important to be kept in mind is that the purpose, or goal, of the situation needs to be relevant to the language which is being used. A definition of authenticity, therefore, should include an awareness of purpose in relation to the context of situation in which language is occurring.

Role of the Participants

One of the basic questions pertaining to authenticity is whether or not language is dialogic or monologic. It is relatively easy to point to the dialogic aspects of speaking activities, but in the case of writing these aspects are often not as obvious. It is said that whereas speech tends to be dialogic, writing is actually monologic (Burns 1990: 38). However, it has also been argued that this point of view fails to take into consideration the social, communicative aspects of written language. The communicative aspects that occur in "oral language learning can occur in reading and writing as well, if reading and writing are understood as fundamentally dialogic." (Peyton & Staton 1991: xv) The perspective that the authors of this paper take is that both spoken and written language are dialogic.

Language is never really one-way. There are degrees of reply and degrees of turn-taking in any language situation. Martin (1992) states that "most writing is expectant of some kind of response, even though it may be some time coming, for example by post or in the form of a critical review" (1992: 512). In fact, written language which requires or expects no reply is unusual. According to Martin, "truly private texts are the exception rather than the rule; some people do keep diaries and journals for themselves, but even these sometimes take the form of letters" (1992: 512-13). This dialogic perspective of language is important in forming a definition of authenticity because in all contexts of situation, whether the language is written or spoken, participants in the language, whether consciously or not, take on certain roles. In evaluating what is done in the language classroom, in terms of authenticity, the dialogic aspects of language need to be taken into consideration, and teachers need to be aware that the development of an understanding of the role that speakers or writers take in any situation is fundamental to authentic communication. Rivers notes that "we need to stimulate interactive situations and activities, if communication, authentic even if elementary, is to take place among our students, or even between teacher and students or students and teacher" (1990: 15). Writing and speaking which is not dialogic, is unlikely to be authentic in any way. Therefore, in order for learners to write or speak successfully, they need to be aware of their roles in the dialogue and the situation itself, and the accompanying linguistic choices which reflect those roles. Authenticity thus includes the notion of dialogic role of the participants in the language.

Conclusion

This paper has set out to discuss the notion of authenticity through a linguistic framework which views language as social activity. Through this investigation

a variety of issues have been highlighted. These include the relationship between language and its context of situation and how this is important in evaluating what is done in the classroom in terms of authenticity, the rich potential of incorporating authentic materials in the classroom, and the value of utilising the language classroom itself as an authentically communicative situation. It is hoped that this discussion into a broader view of authenticity may be incorporated into the teaching/learning situation of the English Language Programme. This paper represents a useful start towards evaluating a language programme for authenticity. As the highlighted issues are incorporated into the overall approach of the English Language Programme and teachers evaluate what they are doing in terms of authenticity, it is hoped that movement towards authentic communication will be maintained, and the development of an enhanced and more genuinely communicative approach to teaching English as a second language will be achieved.

Reference

- Brumfit, C.J. and Johnson, K. 1979. *The Communicative Approach to Language Teaching*, Oxford: Oxford University Press.
- Burns, A. 1990. Focus on Language in the Communicative Classroom. In: Brindley G. (ed.) *The Second Language Curriculum in Action*, Sydney: NCELTR.
- Canale, M. and Swain, M., 1980. "Theoretical Bases of Communicative Approaches to Second Language Teaching and Testing." *Applied Linguistics*, 1 (1): 1-47.
- Carter, R. 1994. *How Aware Should Language Aware Teachers and Learners Be?* Plenary presented at the International Language in Education Conference, Wednesday, 14 December, 1994.
- Chomsky, N. 1965. *Aspects of the Theory of Syntax*, Cambridge, Mass: The M. I. T. Press.
- Crookall, D. and Oxford, R. L. 1990. *Simulation, Gaming and Language Learning*, New York: Newbury House Publisher.
- Ernst, G. 1994. "Talking Circle: Conversation and Negotiation in the ESL Classroom." *TESOL Quarterly*, Volume 28, Number 2, Summer 1994.

- Firth, J.R. 1957. *Papers in Linguistics* 1934-1951, London: Oxford University Press, 1957.
- Galien, P. 1994. The Use of Authentic Videos in a Content-based ESL Curriculum, in *Obirin Review of International Studies*, No. 6.
- Halliday, M.A.K. 1973. *Explorations in the Functions of Language*, (Explorations in Language Study), London: Edward Arnold.
- _____. 1975. *Learning How to Mean: Explorations in the Development of Language*, (Explorations in Language Study), London: Edward Arnold.
- _____. 1978. *Language as Social Semiotic: The Social Interpretation of Language and Meaning*, London: Edward Arnold.
- _____. 1979. Towards a Sociological Semantics, 1979. C.J. Brumfit and K. Johnson, (eds.) In *The Communicative Approach to Language Teaching*, edited by Oxford: Oxford University Press, 1979.
- Halliday, M.A.K. and R. Hasan, 1985. *Language, Context, and Text: Aspects of Language in a Social-semiotic Perspective*, Geelong: Deakin University Press.
- Harmer, J. 1983. *The Practice of English Language Teaching*, London and New York: Longman.
- Hasan, R. 1980. What's Going On? A Dynamic View of Context in Language, *Seventh LACUS Forum*, edited by J. E. Copeland and P. W. Davis, Columbia: Hornbeam Press.
- Hymes, D. H. On *Communicative Competence*, Philadelphia: University of Pennsylvania Press, 1971. Extracts reprinted in C.J. Brumfit & K. Johnson, *The Communicative Approach to Language Teaching*, Oxford: Oxford University Press, 1979.
- _____. 1972. Models of Interactions of language and social life. In *Directions in Sociolinguistics*, eds., J.Gumperz and D. Hymes, New York: Holt Rinehart & Winston, 1972, pp. 35-71.
- _____. 1974. *Foundations in Sociolinguistics: An Ethnographic Approach*, Philadelphia: University of Pennsylvania Press.

- Krashen, S. and Terrell, T. *The Natural Approach: Language Acquisition in the Classroom*, San Francisco: Alemany Press, 1983.
- Labov, W. 1969. *The Logic of Nonstandard English*, (Georgetown Monographs on Language and Linguistics 22) Washington, DC: Georgetown University, Centre for Applied Linguistics. Reprinted in *Language in the Inner City: Studies in Black English Vernacular*, Philadelphia: University of Pennsylvania Press, 1972.
- _____. 1970. *The Structure of Nonstandard English*, Washington, DC,: National Council of Teachers of English and the Centre for Applied Linguistics, 1970.
- _____. 1982. *Sociolinguistic Patterns*, Philadelphia: University of Pennsylvania Press, 1972.
- Martin, J.R. 1992. *English text: System and structure*, Amsterdam: John Benjamins.
- Munby, J. 1978. *Communicative Syllabus Design: A Sociolinguistic Model for Defining the Content of Purpose-Specific Language Programmes*, Cambridge: Cambridge University Press.
- Nunan, D. 1989. *Designing Tasks for the Communicative Classroom*, Cambridge: Cambridge University Press.
- _____. 1988. *Syllabus Design*, Oxford: Oxford University Press.
- Peyton, J. K. and Staton, J. 1991. *Writing Our Lives: Reflections on Dialogue Journal Writing with Adults Learning English*, Englewood Cliffs, New Jersey: Prentice Hall Regents.
- Porter, D. et al. (eds.) 1987. *Methodology in TESOL: A Book of Readings*, New York: Newbury House Publishers.
- Rivers, W. 1981. *Teaching Foreign-Language Skills*, Chicago: University of Chicago Press, Second Edition.

- _____. 1990. *Interaction: The Key To Communication Learning, Keeping and Using Language, Volume 1*, edited by M.A.K. Halliday, J. Gibbons, H. Nicholas, Amsterdam: John Benjamins, 1990, pp. 7-17.
- Scarino, A. et al. 1988a. *Australian Language Levels Guidelines Methods, Resources and Assessment: Book 2*, Commonwealth of Australia: Curriculum Development Centre.
- _____. 1988. *Australian Language Levels Guidelines Syllabus Development and Programming: Book 3*. Commonwealth of Australia: Curriculum Development Centre.
- Seliger, H. and Long, M. 1983. *Classroom-Oriented Research*, Rowley Mass.: Newbury House Publishers.
- Trim, J. L. et al. 1973. *Systems Development in Adult Language Learning*, Council for Cultural Cooperation, Strasbourg: Council of Europe.
- Trim, J. L. 1977. *A European Unit/Credit System for Modern Language Learning by Adults*, , Strasbourg: Council of Europe.
- Wilkins, D. 1976. *Notional Syllabuses*, Oxford: Oxford University Press.
- Wolfson, N. 1989. *Perspectives: Sociolinguistics and TESOL*, New York: Newbury House Publishers.

8

READING AND PROBLEM SOLVING: THE INTERACTION OF INDIVIDUAL DIFFERENCES AND TASK TYPES

Peter Storey

Hong Kong Institute of Education

Introduction

A test may be seen as a means of generating a sample of behaviour upon which to base decisions about general ability. Thus a reading test is an attempt to obtain a sample of reading which is representative of the kind of reading behaviour the candidate would use in the real world. The sample must be representative for the decision to be accurate. We may take a holistic or an analytic approach to this sampling. A holistic approach for example would ask the candidate to perform an authentic task such as using the reading passage as a source for a discursive essay. Reading ability might be measured then in the accuracy with which relevant ideas were selected and synthesised in the essay. An analytical approach on the other hand, might involve analysing the skills involved in reading a passage, and constructing items to measure those skills. This traditional use of text plus test items is still the most common approach to reading assessment. It is however the approach which is most likely to produce inauthentic reading behaviour since it depends on the use of reading test *items* rather than authentic reading *tasks*. Test items may produce representative reading behaviour, but they also produce a wealth of other behaviours. These other behaviours may at best be simply irrelevant to the measurement of reading ability. At worst, they may look like real world reading behaviours, when in fact they are test-specific behaviours, unrepresentative of real world reading. Since the text plus items approach is likely to remain the most common reading assessment method for a long time to come, it is important that we learn more about the characteristics of the items we use.

When selecting items for a reading test we make assumptions about the type of reading behaviour the items are likely to produce. The testing literature contains a number of studies to support these assumptions. Several researchers have looked at different test item types used in the assessment of reading and attempted to characterise the type of behaviour they produce. Cloze tests have been subjected to detailed scrutiny based on differences in results achieved by different subjects and different types and rates of deletion. Short-answer questions, summary and another less common testing

technique - scrambled sentences - have been studied by a number of researchers in the field of psychology. Multiple-choice items, long considered suspect by many in the field of testing because they are open to so many test-taking strategies, have also been the subject of a number of interdisciplinary studies.

The introduction of introspective data has, in the last two decades, brought a new perspective to reading research. A number of studies are available which allow us to get much closer than before to the actual cognitive processes readers employ in constructing meaning from a text. Introspection has been employed not only in the analysis of reading behaviours, but also in the validation of measures used to assess reading ability (Grotjahn 1986, Alderson 1990, Anderson et al 1991). One aspect of this direction in testing research is the identification of test taking behaviours associated with specific test item types (Cohen 1984). This paper attempts to contribute to this body of research, by using the introspective protocols of test-takers to look at five test item types for assessing comprehension in reading - Short-answer Questions, Sentence Completion, Sequencing, Discourse Cloze and Split Sentences.

Multiple-choice

Multiple-choice (MC), though still the major technique employed by the psychometric school of testers has been criticised on a number of counts for general problems such as its openness to cheating and guessing, its dependence on recognition as opposed to production, the difficulty of writing good items and the harmful backwash it may produce (Hughes 1989). Other researchers have looked more closely at the processes employed in tackling MC tests and discovered that subjects employ surface matching and elimination strategies (Dollerup et al 1981), read in a superficial manner, rely on inferences based on the wording of distractors and prior knowledge (Cohen 1984) and read in a fragmented manner directed by the questions (Farr et al 1987). Alderson (1990) suggests that similar behaviours observed in his study are examples of how real reading and reading done in a test-taking context differ from each other.

Cloze Tests

Cloze tests were once heralded as an alternative to MC in that they were much easier to produce, results were consistently reliable, and that they could measure discourse comprehension. Research into the differences

between native speaker and non-native speaker performance and into the effects of different deletion rates and different levels of text difficulty or cloze test results (Alderson, 1979, 1980, Klein-Braley and Raatz 1984) has refuted some of the early claims for cloze. Process investigation shows that far from measuring discourse comprehension, cloze is more a measure of phrase-, or at best, sentence-level reading (Cohen 1984, Markham 1987). Aslanian (1985) used retrospective reports to analyse performance on a MC cloze and found that students were able to use unpredicted clues in the text and MC options to reach correct answers without actually understanding the text either globally or on a micro level.

Short-answer Questions

Short-answer questions are seen as a more reliable means of measuring reading comprehension since they appear to be less open to abuse in the form of test-taking strategies. Short-answer questions bring their own problems however, since test-takers are required to employ composition skills in answering and hence more than pure reading is measured (Hughes 1989, Alderson 1990). Moreover, short-answer questions are the most difficult items to answer perhaps for this very reason (Lewkowicz 1983), though allowing responses in L1 enhances performance (Shohamy 1984). Answering processes for short answer questions are varied, and depend upon the relationship of the question to the text. Textually Explicit, Textually Implicit and Scriptally Implicit questions (Pearson and Johnson 1978) involve quite different answering processes. Goldman and Duran (1988) describe a general procedure for all short-answer question types where the initial Question Encoding stage includes an analysis of the type of question being tackled. The following stages - Memory Search, Search of External Source, Construction and Output of a Solution and Monitoring - vary in degree and character depending on the type of question being answered and also on the proficiency level and prior domain knowledge of the reader.

Scrambled Sentences

Another reading task which has been the subject of study especially in the psychological literature is the unscrambling of scrambled texts. This task is employed as a measure of subjects' ability to utilise text structure. Subjects with a highly developed story grammar are able to recall scrambled stories as well as those which conform to a conventional story structure (Danner 1976, Kintsch et al 1977, Poulsen et al 1979). Extensions of this research to expository texts reach similar conclusions (Taylor and Samuels 1983,

Richgels et al 1987). The general conclusion of these studies is that the ability to restructure scrambled text can be an indication and, by implication, a measure of text structural awareness. This suggestion is given support by anecdotal evidence from Michael Hoey (1983) who describes a similar unscrambling exercise as a Test of Discourse Construction Skills (op cit: 3 ff).

Split Sentences

In contrast to the four test item types described above, little research has been done on the use of split sentences to assess reading. Split sentence "items" are constructed by splitting sentences in half and arranging the second half of each sentence in random order. Candidates then have to choose the best way to complete the first halves of the sentences by choosing from the jumbled second halves. This can be considered as a type of multiple choice task since subjects have to choose final clauses from the options presented to them. As selections are made, naturally the number of choices diminishes. It is postulated that this task would involve recognition of text main ideas in paraphrase form as well as cohesive and syntactic ties between final and initial clauses of the ten sentences. In addition, recognition and use of text structure is postulated to be involved in successful restoration of the summary paragraph.

The Present Study

The findings reported here constitute a secondary analysis which is a spin-off from a test development project involving College of Education students in Hong Kong. This project is described in full elsewhere (Storey 1990, 1992).

Subjects

Twenty-seven native speakers of Cantonese - 26 female and one male - ranging in age from 19 to 24, who were first year students on a two-year course in Grantham College of Education were selected for the study based on their oral fluency in English as measured by a College entry test. The highest scoring first year students who had selected English as one of their two elective subjects were invited to participate in the study. A modest payment was made in order to ensure a higher degree of commitment. The twenty-seven subjects had been exempted from the test when it was

administered, at the beginning of the academic year, to all incoming students to the two-year course - a total of 218 subjects from three different Colleges of Education in Hong Kong.

Materials

The English Diagnostic Reading Test was developed to identify those incoming students who would be unable to cope with the English reading demands of the education component of the College course and would therefore require remedial training in reading skills. Four texts extracted from first year Education textbooks in use in the Colleges formed the basis of the test (Royer et al 1987). Each text was around 500 words in length. Though the texts were taken from different sources, all dealt with different aspects of classroom management.

For each text short-answer questions were written in order to assess components of main idea comprehension identified in the van Dijk and Kintsch (1983) model of reading - paraphrase, recognition of cohesive ties, inferencing, text structural awareness, and understanding of lexical and syntactic meaning. In addition, a summary-paraphrase of the main ideas of each text was drawn up following procedures outlined in Sarig (1989), and used as the basis of different test item types for each text: Sentence Completion, Sequencing, MC Discourse Cloze, and Split Sentences.

Sentence Completion

Royer's (op cit) testing technique - the sentence verification task - was used as the basis for a Sentence Completion task for the first text. Subjects were presented with initial clauses of paraphrased main idea statements and asked to select suitable final clauses from among three options constructed to be either opposite in meaning or irrelevant to the meaning of the text main idea, or to employ varying degrees of surface features matching those in unrelated parts of the text.

Sequencing

For the second text the eight sentences of the summary-paraphrase were scrambled to form a Discourse Construction Test or Sequencing task following suggestions in Hoey (op cit). Subjects were asked to reconstruct

the summary paragraph by reference to the sequence of ideas in the original text.

Discourse Cloze

For text three, the summary paraphrase was used to construct a multiple choice rational deletion Discourse Cloze following suggestion in Bensoussan and Ramraz (1984) and Mauranen (1989). The summary-paraphrase was constructed with explicit markers of discourse structure to highlight the logical, sequential and rhetorical relationships between the main ideas in the text. These markers, together with a few lexical items which acted as cohesive ties, were deleted and presented, together with three distractors of the same form class, as MC options.

Split Sentences

For the final text, the initial clauses of the ten sentences of the summary-paraphrase were presented in the correct sequence, but the final clauses were presented in scrambled order to form a Split Sentences task.

The twenty-seven subjects tackled this battery of tests. The tests differed from the tests given to the main body of 218 students only in that, for the purpose of the introspective study, prompts and reminders were printed in key parts of the test materials to encourage subjects to introspect and retrospect.

Method

The test was administered in a language laboratory, one week after subjects had been given a brief written description of the introspective procedure and a short training session. Subjects were trained in a two-stage introspective process. They were asked first to think aloud while reading the text and tackling the test items, and then to stop after completing each item and describe the method they had used to find the answer. In this way introspective data could be validated by retrospective data as suggested by Ericsson and Simon (1984).

Subjects were free to use Cantonese or English in their protocols, though only two chose to use Cantonese throughout. These two Cantonese protocols were discarded leaving 25 subject protocols for analysis.

The tapes were transcribed and analysed to identify the strategic reading behaviours used. Strategies or moves were identified from the protocols themselves and a taxonomy of reading and test-taking behaviours was drawn up based on descriptions of observed behaviours. This procedure was adopted in order not to distort the data by attempting to fit it into existing strategy descriptions developed on different subject populations. The protocols were recoded after an interval of six months by the same researcher and an average of 87% agreement was found between initial and final coding.

The taxonomy was divided into categories of reading and test-taking strategies and the numbers of each strategy totalled for each item across the 25 subjects. Finally, the total number of strategies recorded in the 12 categories was totalled for each sub-test. This final summary statistic is the subject of the findings reported below.

Results

The following categories of reading and test-taking behaviour were identified in the protocol data:

- I. Using Prior Knowledge
- II. Metacognitive Strategies
- III. Building a Macroproposition
- IV. Inferencing
- V. Using Text Structure
- VI. Using Knowledge of Language and Language Systems
- VII. Recognising Cohesive Ties
- VIII. Matching
- IX. General Problem-Solving
- X. Test-taking Strategies
- XI. Affective Strategies
- XII. Composing Strategies

Because of the small number of observations in the first two categories they were collapsed together leaving eleven categories of behaviour. Composing strategies were only employed in the Short-answer Questions since all the other test-item types were based on recognition rather than production and did not require a written response. The appendix contains a list of the 96 distinct strategies identified across the eleven categories.

The five different test item types showed distinctive patterns of reading behaviour. These are summarised in the table below and discussed in the following section:

	Short- Answer Questions	MC Sentence Completion	Sequenc- ing	MC Summary Cloze	Split Senten- ces
I. Using Prior Knowledge & Metacognition	180 (3.0%)	21 (1.8%)	4 (0.6%)	25 (1.6%)	23 (0.8%)
II. Building a Macroproposition	1102 (18.0%)	391 (33.1%)	83 (13%)	406 (25%)	383 (14%)
III. Inferencing	460 (7.5%)	63 (5.3%)	21 (3.3%)	96 (6.0%)	47 (1.7%)
IV. Using Text Structure	51 (0.8%)	26 (2.2%)	84 (13.2%)	114 (7.0%)	90 (3.3%)
V. Using Language Knowledge	103 (1.7%)	5 (0.4%)	1 (0.1%)	15 (1.0%)	61 (2.2%)
VI. Recognising Cohesive Ties	142 (2.3%)	6 (0.5%)	92 (14.4%)	26 (1.6%)	63 (2.3%)
VII. Matching	348 (5.6%)	178 (15.0%)	129 (20.2%)	107 (6.6%)	752 (28.0%)
VIII. General Problem- Solving	2175 (35.4%)	154 (13.0%)	103 (16.2%)	218 (13.5%)	358 (13.2%)
IX. Test-taking Strategies	633 (10.3%)	305 (26%)	92 (14.4%)	569 (35%)	833 (31.0%)
X. Affective Strategies	283 (4.6%)	32 (2.7%)	29 (4.5%)	44 (2.7%)	95 (3.5%)
XI. Composing Strategies	662 (10.8%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
TOTALS	6139 (100%)	1181 (100%)	638 (100%)	1621 (100%)	2705 (100%)
Average Difficulty (p)	.47	.52	-	.52	.36

Table 1 Number and Percentage of Strategies Observed on Five Measures of Reading Comprehension

The table shows the number of reading and test-taking behaviours observed in each category for the five different test item types together with the percentage of observations in each category, and the average difficulty level for each item type. The Sequencing task was not scored dichotomously, and therefore difficulty estimates are not available. The total number of observed behaviours for each item type is also shown. Short-answer questions were constructed for each of the four texts, while the other testing techniques were only used on one text each. This accounts for the higher number of observations for the Short-answer items. The Split Sentence items were the most difficult and involved the greatest number of reading behaviours. Short-answer questions were the second most difficult type of item, consuming the greatest amount of time and involving the greatest number of revisions and

check backs, all these factors together contributing to the large number of observations. The relationship between average difficulty and average number of behaviours per item is highly significant as calculated by the Student's t-test ($p = 0.0085$). This relationship accords with what intuition would suggest since the reading and test-taking behaviours are problem-solving strategies: the greater the problems encountered with an item, the larger the number of strategies it will involve.

Discussion

The frequency of different types of reading behaviours observed for the different item types can contribute towards our understanding of typical behaviour associated with specific types of test item. In the following discussion specific strategies are referred to using three- or four-letter codes. Explanation of these codes will be found in the Appendix.

Short Answer Items

The greatest proportion of strategies observed for Short-answer questions fall into the category of General Problem-solving (35%), followed by Building a Macroproposition (18%) and Composing (11%). This type of item produces the lowest proportion of Test-taking and Matching strategies and the highest proportion of Inferencing, suggesting that short-answer questions have positive advantages over the summary-based recognition items. However the relatively high proportion of Affective Strategies, which include expressions of frustration and difficulty, highlight the disadvantage noted earlier since these difficulties consist of problems associated with expressing the answer clearly in writing. Subjects whose written answers showed little sign of comprehension, had often demonstrated a clear understanding of the correct answer to the question in their think-alouds.

The high frequency of General Problem-solving strategies for this type of item is a reflection of procedures noted in the protocol data, which are also noted in the psychological literature. What Singer (1990) calls Question Encoding, Question Categorisation and Strategy Selection are called in the strategy taxonomy developed in this study negotiating task demands (NTD) and goal-setting (SGS). When the test-taker encounters the question, the wording of the item must be understood, the targetted concepts identified, the type of question categorised in order that an appropriate goal may be set. Memory Search or Search of External Source (searching the text for the relevant information) may draw a blank revealing that an inappropriate goal

has been set and requiring the test-taker to refer back to the question wording and re-negotiate it. All these strategies fall into the General Problem-solving category. For short-answer items, the negotiation, goal-setting, checking of goals, directed reading of the text (DRT) have to be undertaken afresh for each successive item, since each item may be of a different sub-type of short-answer question, and therefore pose different demands on the candidate. For the other items in the sub-tests, the task usually has to be negotiated once only, since each item in the sub-test poses the same general demands on the candidate. For the summary-based sub-tests candidates read the rubric once, encoding or negotiating the task demands for all the subsequent items in the sub-test.

Underlying the research of which this study forms a small part, is a theoretical model of the reading process based on the van Dijk and Kintsch (1983) framework. Fundamental to this model is the notion that the reader, in processing a text, is building a *macroproposition* or gist summary of the developing propositional meaning in the text as he or she reads. The second category of behaviour is therefore crucial in evaluating the validity of test items. The higher the proportion of behaviours which fall into the category of Building a Macroproposition, the greater the validity of the item type as measured against the theoretical model. For short-answer items 18% of all strategies fall into this category which is a moderately acceptable figure taking into account the fact that the short-answer items for each text were the first ones to be tackled. The fact that test-takers had to deal with the short-answer items before the other item types meant that these items were often tackled before a general understanding of the text had been developed. When tackling the summary based items, test-takers had already read the text through and answered all the short-answer questions, and therefore could use more category II strategies - such as recognising a macroproposition in paraphrase form (MPR), making reference to their own personal gist for the text (UMP), in reaching a solution.

Multiple-choice Sentence Completion

The highest frequency of strategies for the Sentence Completion task falls into the theoretically desirable category of Building a Macroproposition which accounts for a third of all tokens for this sub-test. Reference to the actual numbers of strategies observed reveals however that of 391 tokens in this category, 161 consist of *recognition* of a paraphrased main idea (MPR) rather than *generation* (MPB) of one, which accords with our characterisation of multiple-choice as a recognition-based as opposed to a production-based item type. Nevertheless the sub-test generated a moderately high frequency

of other theoretically desirable strategies in this category including main idea construction processes (CMI/AMI) and reference to an already developed personal gist of the text (UMP). Another predictably high frequency category for this item is Test-taking strategies, which accounts for another 26% of the observed strategies. Of the test-taking strategies, elimination of distractors (ELN - 205 tokens) together with rehearsal (RSR - 70 tokens) account for 90% of observations. Rehearsal consists of trying out various combinations of stem and options to find a suitable combination. This would not seem to be a desirable reading behaviour in the sense that it is not the kind of reading likely to be used in real world reading tasks. The high frequency of Matching strategies in this sub-test is another indication of undesirable reading behaviours, but this indication is tempered somewhat by reference to the actual strategies observed which reveals that the larger number of these were meaning-based *semantic matching* (SMM) strategies rather than matching of surface features (SF), and that the surface matching strategies were associated with incorrect choices.

The general pattern of behaviour for this item type consisted of eliminating distractors, trying out combinations of stem and MC option, and scanning the text to locate a proposition matching that of one of the MC options (SSMM). The protocol data show that these components of the general answering procedure occurred in two different sequences. Some subjects would read the stem and all the options making a provisional selection based on their already developed macroproposition for the main text. This provisional selection would then form the basis for a text search to locate the matching proposition in the text, in order to confirm their final selection of option. Others would read and encode the stem alone, then use it as the basis for a text search to locate the semantically matching source of the paraphrase. These subjects would then re-read the section of the main text *following* the located source and use their interpretation of that as the criteria for evaluating the options. Therefore, elimination and rehearsal may often be complex strategies which do involve some processing of the paraphrase paragraph and the main text. The association of these test-taking strategies with text search and semantic matching lends them greater respectability.

Sequencing

The van Dijk and Kintsch (1983) model of the reading process acknowledges the importance of a knowledge of text structure in the reading process. The Sequencing task is included in the test because it is hypothesised to involve this component of text processing ability. The frequencies of strategy use observed in the protocol data for this sub-test

confirm this hypothesis to a certain extent. The strategy coded as using text structure (UTS) accounts for a greater proportion of observed behaviours for this test item type than for any of the others. However only 13% of observed strategies for this sub-test fall into this category which is not as high as would be expected. The largest frequency of strategies for the sub-test falls into the undesirable category of Matching. Moreover, the most frequent type of matching observed is based on shared surface features (SF) With 71 tokens, this is the most frequent strategy of all, though the more desirable semantic matching, which comes second with 50 tokens, is also an important strategy in the sub-test. The high frequency of surface matching is partly accounted for by the recognition of cohesive ties (RCTR), since cohesion is established in the summary-paraphrase, by the repetition of noun phrases. Subjects who recognised cohesion by seeing connections between sentences on the basis of a shared noun phrase were using both the surface matching and recognition of cohesion strategies. Thus recognition of cohesion, which would normally be considered a desirable discourse strategy, is revealed here as a variety of the theoretically undesirable surface matching.

Subjects tackling this sub-test adopted two general approaches to the task. One consisted of a search for connections between the sentences to be sequenced, the other consisted of a search for sequences of thematically related propositions. The majority of subjects used elements of both general approaches, though some subjects confined themselves to one or the other. The search for *connections* led subjects to look for repeated noun phrases (STSF), lexical substitutes employed as cohesive ties (RCTL), and the antecedents of anaphoric pronouns (RCTP). It was in the first of these behaviours that the undesirable reading behaviour was observed. A text with fewer repeated noun phrases, with lexical substitution used instead of repetition, might produce a lower proportion of surface matching and a major problem with this item type might be considerably reduced.

The search for *sequences* led subjects to look below the surface at the propositional content of the sentences and to identify rhetorical features such as introduction and conclusion, topic and comment. Two secondary approaches were observed in this seemingly higher level approach. Some subjects relied on their recall of text content or the macroproposition they had internalised for the main text (UMP). Such subjects would recall the text main ideas in propositional form (RTM) and search for semantically matching propositions in the sentences to be sequenced. Other subjects preferred to make explicit reference back to the main text and hence employed a combination of surface and semantic matching to identify the source which had been paraphrased in the sentences to be sequenced.

A successful general strategy for this sub-test consisted in, first, a search for the opening sentence of the paragraph. Having identified the opening sentence candidates would refer back to the main text to find the original source paraphrased in this opening sentence. Having located the source of the paraphrase, subjects would re-read the *following* section of the main text reminding themselves of its content, and then search for matching propositional content in the remaining summary sentences. This process might involve *elimination* to reject or select candidate sentences, *semantic* and occasionally *surface* matching to identify paraphrases of text ideas, the use of text structure, evaluation of the sequence of information (INS), and recognition of cohesive ties to establish the feasibility of the connection made between the two summary sentences. This procedure would be repeated for each successive sentence until the complete paragraph had been reconstructed, and then a *resolution* stage would be entered in which the completed paragraph would be evaluated for coherence.

Multiple-choice Summary Cloze

The use of multiple-choice options in cloze sub-test inevitably led subjects to use elimination as a major strategy for restoring deletions. Together with the strategy which in this study is called rehearsal (RSR - 198 tokens), elimination (275 tokens) accounts for the apparent dominance of the ninth category of Test-taking strategies in the profile of test-taking behaviour for this sub-test. More desirable reading strategies are also well-represented in Table 1 above, however, which shows that 406 tokens of macropropositional strategies were observed, amounting to 25% of all strategies for this part. These include a majority of *recognition* processes (MPR - 143 tokens), but also include references to already developed personal gists of the main text (UMP - 114 tokens) and a high proportion of automatised main idea construction processes (AMI - 134 tokens). Test-taking and Macroproposition Building strategies account for 60% of all strategies observed for this part. Among the other 40% of strategies are indications that this sub-test produced a generally desirable style of reading. These indications include a relatively high proportion of inferencing (INF) - at 6% second only to Short-answer Questions - and Use of Text Structure - at 7% second only to Sequencing - and low frequencies of Matching and Affective strategies. The low frequency of strategies in the category of recognising Cohesive Ties however, indicates a general failure of subjects to see the task as one of paragraph building rather than sentence-level gap-filling.

This sub-test was designed to measure higher level discourse comprehension processes including utilisation of text structure, information

sequence, discourse markers and cohesive ties. Subjects introspecting on the items in the sub-test reported using these strategies, but other, less desirable test-taking strategies were produced in larger numbers. Elimination of options, the most common strategy observed, was done largely on the basis of localised clues and a general feeling for what sounded right rather than on a clearly developed analytical awareness of the rhetorical relationships being reconstructed in the restoration of deletions. Recognition of cohesion very often turned out in practice to be identification of a repeated noun phrase - a variety of surface matching. A relatively high frequency of guessing was observed in the protocols.

A common approach to the restoration of deleted discourse markers involved a kind of *decomposition* in which the relationship between arguments was analysed in order to select the appropriate marker to connect them. This appeared, in the think aloud protocols to be a high level cognitive process. Candidates often engaged in complex elaboration (ELB) of the text arguments in order to get at the underlying rhetorical relationships. However, decomposition was often localised in the sense that reference was not made back to the main text or use made of a macropropositional gist for that text. Rather, the decomposition often related to the arguments expressed in the summary paragraph itself, and was often confined to the sentence level. The answer to the oft-repeated question as to whether cloze measures constraints across sentences (Chihara et al 1977, Alderson 1979, Bachman 1982, 1985, Porter 1983, Chavez-Oller et al 1985) would appear to be that it has the capacity to do so, but that candidates often rely on lower level routes to the correct answer.

On the whole this was a cognitively demanding task despite its moderate average difficulty level ($p = .52$). It is interesting that the level of cognitive difficulty is often not apparent in subjects' scores, but is revealed in the protocol data. Expressions of difficulty and frustration, complex answering processes with resort to a large number of different strategies including compensatory test-taking strategies, miscues of unfamiliar items - all these clues are available in the protocol data and point to the difficulty of the task.

Split Sentences

The Split Sentences task generated a high proportion of Test-taking and Matching strategies. Together these two categories of behaviour accounted for nearly 60% of all strategies observed on the sub-test. Building a Macroproposition accounts for 14% of strategies observed, with General Problem-solving accounting for a further 13%, leaving another 13% of

strategies spread over the remaining six categories. The high frequency of Test-taking strategies is accounted for largely by strategies associated with elimination, and rehearsal, together with the deferral of difficult items until later (DTI). The Macroproposition strategies consisted of subjects making reference to an already developed gist for the main text (UMP) and recognising paraphrases of text main ideas in the summary paragraph (MPR). The General Problem-solving strategies consisted largely of the giving, revising and confirming of provisional responses (GPR/RGR/CGR) - which is partially associated with the deferral strategy recorded in the Test-taking category. In general then, the Split Sentence task appears to generate a large range of strategic processing behaviours which appear to be quite unrelated to real reading behaviours. Although inauthentic reading behaviours were generated, an argument can nevertheless be made for the validity of the task on the basis of the type of cognitive processing it appears to generate.

Three possible levels of processing were identified in tackling the task. The introspective data revealed that some subjects were able to exploit all three levels in making an initial response and then verifying it, while others worked on only one or two of the available levels. The three levels were semantico-grammatical, macropropositional and paraphrase. Thus, semantico-grammatical criteria might be used to match an initial sentence stem with a final clause from the available options. This might then be verified by elaboration in which the subject makes explicit the relationship of the completed sentence to the main ideas of the main text. The selection might then be confirmed by reference back to the main text to locate the semantically matching propositions.

The most common strategy observed was elimination followed by rehearsal, scanning and semantic matching. reference to macropropositions only came sixth in order of frequency followed closely by General Problem-solving and Test-taking strategies. Elimination is not considered an example of real reading, and on this basis we might conclude that this item type produces a specialised type of reading only loosely related to authentic reading in the real world. Examination of the protocol data do suggest that elimination might involve more desirable reading and cognitive processes than we might at first assume, however. In the data we observe the following procedure. First the task demands are set and appropriate goals negotiated in a familiarisation stage involving reading through the sentence stems and options. Then, a typical approach would involve searching for an appropriate completion for the stems in serial order. This would involve setting search criteria based on examining the grammatical structure and semantic content of each stem in turn, as well as referring this local semantic content to a

global semantic content developed in the subject's macroproposition for the text. Eliminating options then would be done on the basis of multiple criteria.

Elimination is seen to be a relatively complex cognitive task especially when a number of options have to be evaluated for their degree of fit to these criteria. For the first item, for example, the grammatical and both the local and global semantic content of the stem need to be held in short-term memory while ten options are evaluated in turn. As subjects proceed through the sub-test and the number of available options are reduced, the cognitive demands may be reduced also, but this assumes that options already selected are not re-considered in searching for completions for later stems. The high number of provisional selections and revisions observed in the data show that this is not the case. Already selected options are repeatedly reconsidered in searching for completions to other stems.

Elimination is often observed in combination with the deferral strategy in the data, a tendency which reduces its acceptability somewhat. Subjects tackle the easier items first and defer more difficult ones until later. This has the inevitable effect of localising the selection criteria to a certain extent, since the argument in the summary paragraph is treated in a fragmented fashion rather than being followed through in a serial manner. This means that discourse is neglected in favour of an emphasis on local semantico-grammatical connections. Some subjects attempted to think (aloud) through their choices in an elaboration strategy (ELB, ELB/MPB). Elaborating on a selection counteracts this localisation to a degree since elaboration consists of accommodating the completed sentence to the subject's macroproposition for the text. In cases where incorrect selections have been made, elaboration leads to some manufactured interpretations (MTI). In the case of more difficult stem-option matching tasks, the elaboration may be neglected, perhaps due to the limited supply of attention available having been used up by the cognitive demands of the task itself. Another reason for the omission of the elaboration stage, which we can infer from the data, is that subjects have made a selection which they cannot justify or accommodate easily since it is not a correct one, and so they proceed in somewhat embarrassed silence to the next stage. In general, for this task as for the others in the test as a whole, when the difficulty of the task increases, the explicitness of the protocols decreases.

Conclusions

According to Sarig (1987) "... most of the reading comprehension process consists of the reader's use of a unique, personal combination of moves

which characterizes them as individuals." The introspective data for the present study supports this view of reading as an idiosyncratic process involving unique combinations of strategies. This places severe limitations on what we can learn about the reading and test-taking process in general from a study of the average number of strategies employed in specific test tasks by a group of 25 idiosyncratic individuals. Nevertheless some interesting general characteristics of selected test item types have been tentatively identified in the present study and might serve as a useful starting point for further research.

A crude measure of the authenticity of the reading produced by the different testing techniques would be the proportion of reading behaviours produced which fall into the first six categories of strategies, since these are the behavioural components of a theoretically sound model of the reading process. None of the test item types examined in this study generated *only* authentic reading, each of them included a proportion of test-specific reading behaviour. A crude measure of the acceptability of test item types would be to look at the proportion of Matching and Test-taking behaviours they generated. The higher the proportion of strategies falling into these categories the more suspect the technique is. Thus Sequencing and Split Sentences are suspect due to their unacceptably high proportions of Matching strategies; Split Sentences and the Multiple-choice item types are suspect due to their high proportions of Test-taking strategies.

The analysis of subjects' behaviour on the tasks has, however, raised the possibility that strategies usually considered undesirable may have some features which make them useful as measures - albeit indirect measures - of reading ability. Semantic matching may be a recognition of a main idea in paraphrase form, while surface matching may be a recognition of a cohesive tie established by repetition. Elimination may involve the use of multiple evaluation criteria in a complex cognitive behavioural pattern. A general conclusion from the study would appear to be therefore, that it is difficult to generate and then measure real reading behaviour using traditional text and test items methods. Perhaps we should direct our efforts to discovering more about the cognitive processes involved in tackling indirect measures of reading so that we can make more confident judgments about the relationships between test-specific reading and real reading ability.

References

Alderson, J. C. (1979) The Cloze Procedure and Proficiency in English as a Foreign Language. *TESOL Quarterly*, Vol. 13, No. 219- 227.

- Alderson, J. C. (1980) Native and Nonnative Speaker Performance on Cloze Tests. *Language Learning*, Vol 30, No. 1, 59-76.
- Alderson, J. C. (1990) Testing Reading Comprehension Skills (Part Two) *Reading in a Foreign Language*, Vol. 7, No. 1, 465-504.
- Anderson, N.J., Bachman, L., Perkis, K. and Cohen, A. (1991) An Exploratory Study in the Construct Validity of a Reading Comprehension Test: Triangulation of Data Sources. *Language Testing*, Vol. 8, No. 1, 41-46.
- Aslanian, Y. (1985) Investigating the Reading Problems of ESL Students. *English Language Teaching Journal*, Vol. 39 No. 1, 20-27.
- Bensoussan, M. and Ramraz, R. (1984) Testing EFL Reading Comprehension Using a Multiple-Choice Rational Cloze. *Modern Language Journal*, Vol. 68, No. 3, 230-239.
- Chavez-Oller, M.A., Chihara, T., Weaver, K.A. and Oller, J.W. jr. (1985) When are Cloze Items Sensitive to Constraints across Sentences? *Language Learning*, Vol. 35, No. 2, 181-206.
- Chihara, T., Oller, J., Weaver, K. and Chavez-Oller, M.A. (1977) Are Cloze Tests Sensitive to Constraints across Sentences? *Language Learning*, Vol. 27, No. 1, 63-73.
- Cohen, A. D. (1984) On Taking Language Tests: What the Students Report. *Language Testing*, Vol. No. 70-81.
- Danner, F. (1976) Children's Understanding of Intersentence Organization in the Recall of Short Descriptive Passages. *Journal of Educational Psychology*, Vol. 68, No. 2, 174-183.
- Dollerup, C., Glahn, E. and Rosenberg-Hansen, C. (1981) Reading Strategies and Test-solving Techniques in an EFL Reading Comprehension Test - a Preliminary Report. *Journal of Applied Language Study*, Vol. 1, 93-99.
- Ericsson, K. A. and Simon, H. A. (1984) *Protocol Analysis: Verbal Reports as Data*. Cambridge Mass.: MIT Press.

- Goldman, S. R. and Duran, R. P. (1988) Answering Questions from Oceanography Texts: Learner, Task, and Text Characteristics. *Discourse Processes*, Vol. 11, 373-412.
- Grotjahn, R. (1986) Test Validation and Cognitive Psychology: Some Methodological considerations. *Language Testing* Vol. 3, No. 2, 159-185.
- Farr, R., Pritchard, R. and Smitten, B. (1987) A Description of What Happens When an Examinee Takes a Multiple-Choice Reading Comprehension Test. *Journal of Educational Measurement*, Vol. 27, No. 3, 209-226.
- Hoey, M. (1983) *On the Surface of Discourse* London: George Allen & Unwin Ltd.
- Hughes, A. H. (1989) *Testing for Language Teachers*. Cambridge: Cambridge University Press.
- Kintsch, W., Mandel, T. S. and Kozminsky, E. (1977) Summarizing Scrambled Stories. *Memory and Cognition*, Vol. 5, No. 5, 547-552.
- Klein-Braley, C. and Raatz, U. (1984) A Survey of Research on the C-Test. *Language Testing*, Vol. 1, 134-46.
- Lewkowicz, J. (1983) Method Effect in Testing Reading Comprehension: A Comparison of Three Methods. Unpublished MA dissertation, University of Lancaster.
- Markham, P. L. (1987) Rational Deletion Cloze Processing Strategies: ESL and Native English. *System*, Vol. 15, No. 3, 303-311.
- Mauranen, A. (1989) Can Gaps Measure Comprehension? Modifications of Cloze as Tests of Reading. In C. Lauren, and M. Nordman, (Eds.) *Special Language*. 337-346. Clevedon: Multilingual Matters
- Pearson, P. D. and Johnson, D. D. (1978) *Teaching Reading Comprehension* New York: Holt, Rinehart & Winston.
- Porter, D. (1983) Assessing communication proficiency. In Johnson and D. Porter (eds.) *Perspectives in Communicative Language Teaching*. 189 - 204. London: Academic Press.

- Poulsen, D., Kintsch, E. and Kintsch, W. (1979) Children's Comprehension and Memory for Stories. *Journal of Experimental Child Psychology*, Vol 28, 379-403.
- Richgels, D. J., McGee, L. M., Lomax, R. G. and Sheard, C. (1987) Awareness of Four Text Structures: Effects on Recall of Expository Prose. *Reading Research Quarterly*, Vol. 22, No. 2, 177-196.
- Royer, J. M., Abranovic, W. A. and Sinatra, G. (1987) Using Entering Reading Comprehension Performance as a Predictor of Performance in College Classes. *Journal of Educational Psychology*, Vol. 79, No. 1, 19-26.
- Sarig, G. (1987) High-Level Reading in the First and in the Foreign Language: Some Comparative Process Data. In J. Devine, P. L. Carrell, and D. E. Eskey (Eds.) *Research in Reading in English as a Second Language*, 105-120. Washington, D.C. : TESOL.
- Sarig, G. (1989) Testing Meaning Construction: Can we do it Fairly? *Language Testing*, Vol. No. 77-94.
- Shohamy, E. (1984) Does the Testing Method Make a Difference? The Case of Reading Comprehension. *Language Testing*, Vol. 1, No. 2, 147-70.
- Singer, M. (1990) Answering Questions about Discourse. *Discourse Processes*, Vol. 13, 261-277.
- Storey, P. (1990) A process approach to test validation. In Verner Bickley (Ed.), *Language Use, Language Teaching and the Curriculum*. Hong Kong Education Department.
- Storey, P. (1992) What does this text mean? Assessing the limits of limited reading comprehension. In Verner Bickley (Ed.), *Quill and Quilt*. Hong Kong: Education Department.
- Taylor, B. M. and Samuels, S. J. (1983) Children's Use of Text Structure in Recall of Expository Material. *American Educational Research Journal* Vol. 20, No. 4, 517-528
- van Dijk, T. A. and Kintsch, W. (1983) *Strategies of Discourse Comprehension*. New York: Academic Press.

Zuck, L. V. and Zuck, J. G. (1984) The Main Idea: Specialist and Non-Specialist Judgments. In A.K. Pugh, and J. M. Ulijn, *Reading for Professional Purposes*. London: Heinemann Educational Books.

Appendix

Description of Reading Behaviours Identified in Subject Protocols.

I. Using prior Knowledge and Metacognitive Strategies

CODE	DESCRIPTION
RPK	Relating text information to prior knowledge of the topic area as a method of assimilating or comprehending the information.
PTM	Personalising text meanings as an aid to assimilation.
PCT	Providing a personalised commentary on the text - agreeing, disagreeing etc...
PQT	Posing questions of the text.
MCN	Monitoring one's level of comprehension.
MPN	Making predictions of text content based on an understanding of what has gone before.

II. Building a Macroproposition - collapsing the propositional content of the text into a macroproposition discarding the actual wording and holding the meaning or propositional content of the main ideas in memory store.

MPB	Forming the gist of the text as it is read, and holding that gist in memory as new input is taken in.
ELB/MPB	Elaborating on text meanings and forming a mental representation in the process.
MPB/PAR	Paraphrasing part of the text and incidentally forming a gist of the text main ideas.
MPR	Recognising the macropropositional gist of the text in a paraphrase of the text main ideas which avoids use of surface structures used in original text.
UMP	Referring to a preconceived macropropositional gist in solving problems related to the text.
CMI	Crunching - pausing to allow automatic MI construction processes to work on already gathered input.
AMI	Automatic MI construction.
GIS	Gathering ideas together and selecting from them to compose a MI statement or summary.
RTM	Recall of text meanings.
RTW	Recall of precise text wording.

III. Inference - recognising that what is stated in the text has some implication which is not explicit. Making the inference necessary to fill in the gaps in what would be a fully explicit account.

INF	Making the inference necessary to connect two propositions to form a complete argument, when they are not explicitly linked in the text.
INFI	Making explicit an implication which is only implicit in the text.
INFL	Making explicit a logical deduction which is only implicit in the text.
ELB	Elaborating an interpretation in order to solve a problem related to the text.

IV. Using Text and Information Structure as a basis for the solution of problem.

DEF	Recognising entities which fulfill the conditions stated in a definition and thus qualify as examples of the entity defined.
UDM	Using one's knowledge of the function of discourse markers to establish the accuracy of a decision or choice in problem solving.
UTS	Using one's conceptualisation of the structure of the text as the basis for a solution to a text/test task.
INS	Using normal order of given/new information (as a way of sequencing sentences to form a summary of the text).
DSP	Distinguishing subordinate from primary information.
MCT	Making connections between different parts of the text.
DSC	Using the text as a whole to provide an answer rather than relying on shared chronology of text and test items.

V. Using knowledge of language and language systems as the basis of a judgment.

UGK	Using grammatical knowledge as the basis of a solution to a test/text task.
WLI	Subject makes a decision based upon an interpretation of a single word in the text
REC	Reconstructing an elliptical clause.
GLC	Guessing the meaning of lexical item from surrounding context in which it is embedded.
UWF	Using clues from the formation of a lexical item as a means of guess at its meaning.
UPC	Using clues from the punctuation of a text as a means of inferring meaning when faced with a problem.
ULM	Understanding the literal meaning of the text.
SMM/COL	Using knowledge of collocation as a means of selecting or constructing solutions to a test task.

VI. Cohesive Ties - recognition of methods by which Main Ideas are restated in the text thus reinforcing the MI status of the idea in the macropropositional gist.

RCTP	Recognition of cohesive ties established by pronouns referring anaphorically to antecedent NPs
RCTL	Recognition of cohesive ties established by use of a semantically equivalent or closely related lexical item substituting for an antecedent NP
RCTR	Recognition of cohesive ties established by repetition of a lexical item.
CCTL	Creating a cohesive tie established by use of a semantically equivalent or closely related lexical item substituting for an antecedent NP, or by pronouns referring anaphorically to antecedent NPs.

VII. Matching - noting that the same lexical items/morphemes or propositions are used in the text and in test items.

SF	Using the fact that a lexical item is shared between text and test item or stem and option as the sole basis for the solution to a test task.
SMM	Recognising that individual meanings or propositions are repeated in text and test items without referring to the general context.
SSMM	Searching the text for a matching exponent of a propositional value held in STM.
STSF	Searching the text for a surface match - to locate the source of an answer (WCI), to verify that remainder of text does not contain an answer (RRT), in order to locate relevant section of text for more intensive reading.

VIII. General Problem-Solving Strategies

SGS	Setting goals in planning a test taking tactic establishing carefully what the task demands are and directing behaviour accordingly.
CTD	Checking task demands - over and above the normal reading of the test items, and as an attempt to deal with ambiguity.
NTD	Negotiating task demands, elaborating in order to be clear about what is actually asked for in task rubric.
WCI	Widening the available context by reading further back or further forward in a text in order to resolve a problem.
RFI	Making a judgment based upon an intuitive impression of what "sounds right".
RSR	Rehearsal - juxtapositioning sentences/stems/options in order to find a combination which sounds right.
GPR	Giving a provisional response, a response which is considered provisional because candidate feels uncertain about its accuracy and feels that remainder of text/test items may give further clues to the correct answer.
RGR	Revising a response given earlier to a test task.
CGR	Confirming a response given earlier to a test task.

IX. Test-taking strategies.

CDR	Faced with a test task which is problematic, copying words or phrases from the text, deleting others, rearranging the order of the copied words and phrases hoping to gain some of the marks available for the item.
GSS	Using guesswork or unreasoned speculation based on random factors or chance to select an answer.
ESI	Examining the surrounding test items in order to help in solving a test/text related problem.
ASC	Assuming shared chronology between text and test items.
ELN	Eliminating distractors.
QDD	Questioning or examining the distribution of letter codes - ABCD in a MC as a means of evaluating ones own response.
UCI	Using clues in the wording of the test item itself to infer that a particular answer is correct
LAS	Local answering strategy: subject selects a phrase from the text which is in close proximity to a phrase targeted in the test item, when the item is not related to the task demands of the item
ELB/CDR	Subject elaborates an interpretation of text using combination of text wording and own words.
DRT	Directing the reading of the text towards answering of the test items eg reading the test items first in order to establish a purpose for reading the text itself.
SRT	Selective reading of the text in order to solve the problems posed by test items/tasks.
RRT	Re-reads the text to aid comprehension.
DTI	Deferring an item for which no answer can be found until later in a test taking session - doing the easier ones first in the hope that the more difficult ones will be more easily tackled after test-bias effect.
RDI	Reappraising a difficult item in the light of knowledge/familiarity gained from tackling the easier items first.
TFL	Translating text or test items into L1.
MTI	Manipulating a text to make it fit in with a pre-conceived interpretation.
JCC	Jumping to a conclusion based upon a cursory reading of the text.

X. Affective Strategies

SIR	Stating inability to respond.
EIP	Expressing inability to carry out the test task.
SLU	Stating lack of understanding - note must be followed by some other strategy since it may be an expression of frustration, or a means of girding one's loins.
EUR	Expressing uncertainty about the accuracy of a response.
EDR	Expressing dissatisfaction with response.
ECR	Expressing confidence in the accuracy of a response.
GOL	Self-encouragement
QTD	Questioning the validity or purpose of the task demands insofar as it boosts confidence or acts as a means of distancing oneself from one's own response.
ATP	Expressing anxiety about time pressure.
EDF	Expressing dissatisfaction or frustration with test/task demands
CDT	Commenting on the difficulty of a text or task
EPP	Expressing physical problems such as tiredness or hunger

XI. Composing Strategies.

CWR	Composing a written response.
RSM	Retrieving words of equivalent/semantic propositional value from own interlanguage store.
CWR/CDR	Composing a response by the copy/delete strategy.
CWR/RSM	Composing a response using own words.
SOW	Searching for own words to express a proposition obtained by reading.
POW	Planning a written response.
RPKP	Utilising a familiar lexical structure (formal lexical schema) in composing a response.

9

"WILLINGNESS TO COMMUNICATE" AMONG HONG KONG STUDENTS

Barry Asker

English Department, Lingnan College

Introduction

Over the years there has been much anecdotal evidence suggesting that Hong Kong students are orally weaker than their peers in other Asian countries and that students evince considerable unwillingness to communicate in English in academic contexts. Industry leaders continue to make similar complaints, arguing that the resulting poor quality of English speech has a negative effect on Hong Kong's ability to perform in the international commercial arena.

Such evidence has played its part in the establishment of an oral component in the Use of English examination. The hope is that by putting oral performance on the test agenda, more time and effort will be put into oral competence. However, there is as yet little hard objective evidence to suggest that in fact Hong Kong students are excessively reticent and even less so show that such lack of willingness to speak up in English implies an equal inability so to do.

This present paper, then, seeks to measure willingness to communicate among Hong Kong students and to establish whether the "Willingness to Communicate" instrument (McCroskey, J.C., & Richmond, V.P., 1987) is a reliable and efficient way to identify orally reticent students (in great numbers) and so create opportunities for extra training or more directed class groupings.

Oral Communication Constructs

Communication research in America has developed a number of approaches to oral communication. These have covered a number of types of general oral behaviours such as shyness, reticence, stage fright, communication apprehension, unwillingness to communicate, speech anxiety and others. The purpose of this approach is better to understand the factors which contribute to ineffective oral communication (especially among ESL students) and to seek ways to enable them to overcome perceived difficulties. /

Research has indicated three distinct but related kinds of communication construct. The first focuses on *anxiety or apprehension* about communicating

and includes stage fright, communication apprehension and speech anxiety. The second group of constructs concerns itself with frequency of talking: verbal activity, vocal activity and talkativeness. The third is a variation on the second and centres on the approach or avoidance of communication, including reticence and unwillingness to communicate. It is considered that these three constructs together comprise a broad approach to identifying and distinguishing certain oral communication behaviour features.

This research has led to insights of great variety into various aspects of oral communications, such as in bilingualism (McCroskey, Fayer & Richmond, 1985), identifying academically at-risk students (Chesebro, McCroskey, Atwater et al, 1992), and the effectiveness of remedial programmes in overcoming oral communication weaknesses (Kelly & Keaten, 1992), to mention but a few. There has also been considerable work in assessing the relationship between anxiety and language learning contexts (MacIntyre & Gardner, 1991), including the relationship between apprehension in communicating and foreign language oral proficiency (Young, 1986). It has been suggested that students with lower levels of anxiety tend to learn better (MacIntyre & Gardner, 1989; Ericson & Gardner, 1992), are more willing to volunteer answers in language classes (Ely, 1986), and are more active socially with the target language group (Clement, 1987).

In short, there is much in the literature which suggests that anxiety and reticence adversely affect performance in language acquisition. This is of obvious importance in the teaching of foreign languages in that if ways of reducing anxiety and overcoming reticence can be found, then so-called "shy" students will be enabled to make faster progress in learning a language. Of course, it is scarcely surprising that students are more reticent in communicating in a language which they know only imperfectly, at best.

If all this is true for the classroom learning environment, it is equally true of oral testing contexts. Increasingly, testing of English (and other languages) has placed an emphasis on oral communication skills and students are required to show that they can speak to an examiner. This follows the dominant concept of communicative, functional performance as opposed to more formal grammar-based instruments of assessment. This pressure on students to perform orally may be seen to disadvantage "shy" students, i.e. those suffering from high apprehension and reticence, and there is a case for identifying such students and considering remedial measures to enable them to cope.

The recent introduction of an oral component in the Hong Kong Use of English examination, comprising 20% of the total mark, brings the problem of shy or reticent students to the fore. It is clear that students who suffer from a

reticent trait characteristic are likely to fare worse than their more vocally confident peers. It should be noted that a good grade in the Use of English examination is a key factor in securing a place at one of Hong Kong's tertiary institutions.

Rationale of this Study

Against this background, the present paper seeks to use the tools of communication research to begin the process of discovering Hong Kong norms of reticence. I wish to find out if Hong Kong students are more or less reticent than other comparable nationality groups and whether this reticence finds a correlation in test performance results.

The research presented here is a study of one particular instrument, the "Willingness to Communicate" (WTC) probability-estimate scale, conceptualised by McCroskey & Richmond (1987). This instrument is a 20-item self-report procedure devised originally with American native speakers in mind. The scale seeks to measure a student's "predisposition toward approaching or avoiding the initiation of communication" and clearly fits the third kind of construct referred to above (*approach/avoidance* concept).

The first question to ask is whether the instrument in its original form would work with Hong Kong ESL students. Would they understand how they were to respond to the questionnaire? Or would the instrument reflect a too strong American bias? If the instrument failed at this level, it would be unlikely to provide a platform for further large-scale investigations into student reticence in Hong Kong. At the same time, the sample of first year undergraduate students at Lingnan College provide initial data on rates of reticence, under several distinct categories, which could be compared with those of other nationality groups.

Description of the Instrument

The "Willingness to Communicate" scale has 20 items and is a self-report measure. Of the 20 items, 8 are fillers and the remaining 12 are scored. The scale is sub-divided, when scored, into seven subscores. Four of these are based on types of communication context (*dyad, group, meeting and public speaking*), and three are based on types of receivers (*strangers, acquaintances and friends*). In short, the instrument crosses three types of receiver with four kinds of communication context so as to provide a representative, though perhaps not exhaustive, matrix of communication situations.

In the scoring, the responses to each sub-category are added together and expressed as a percentage figure. For example, add 4,9,&12 (dyad) together; add 8,15,&19 (group); 6,11,&17 (meeting); 3,14,&20 (public speaking). Divide each sub-total by 3 to arrive at a percentage score. Then, add 3,8,12,&17 (stranger); 4,11,15,&20 (Acquaintance); and 6,9,14,&19 (Friend). The total Willingness to Communicate score is calculated by adding the Stranger, Acquaintance, and Friend sub-scores together and dividing by three.

Figure 1

Willingness to Communicate Scale (WTC)

DIRECTIONS: *Below are twenty situations in which a person might choose to communicate or not to communicate. Presume you have completely free choice. Indicate in the space at the left what percent of the time you would choose to communicate in each situation. 0 = never, 100 = always*

- _____ 1. *Talk with a service station attendant.
- _____ 2. *Talk with a physician.
- _____ 3. Present a talk to a group of strangers.
- _____ 4. Talk with an acquaintance while standing in line.
- _____ 5. *Talk with a salesperson in a store.
- _____ 6. Talk in a large meeting of friends.
- _____ 7. *Talk with a police officer.
- _____ 8. Talk in a small group of strangers.
- _____ 9. Talk with a friend while standing in line.
- _____ 10. *Talk with a waiter/waitress in a restaurant.
- _____ 11. Talk in a large meeting of acquaintances.
- _____ 12. Talk with a stranger while standing in line.
- _____ 13. *Talk with a secretary.
- _____ 14. Present a talk to a group of friends.
- _____ 15. Talk in a small group of acquaintances.
- _____ 16. *Talk with a garbage collector.
- _____ 17. Talk in a large meeting of strangers.
- _____ 18. *Talk with a spouse (or girl/boy friend).
- _____ 19. Talk in a small group of friends.
- _____ 20. Present a talk to a small group of acquaintances.

* Filler item (Not shown on student questionnaire)

Administering the Instrument: Pilot Project, Spring 1994

Figure 1 presents the WTC in its original form as it was administered to 124 Lingnan Students. The teaching staff administering the instrument remarked unanimously that some students expressed difficulty in understanding the questionnaire and needed help. McCroskey (1989) notes that although some difficulties were recorded in non-native speakers in other trials, once a few explanatory points had been made, most students were comfortable with the instrument. The Lingnan College experience would suggest that this understates student confusion. Simple word changes could help in this respect. For example, the word "physician" can be exchanged for "doctor", and "service station attendant" could be replaced with "petrol station attendant".

Greater difficulties presented with respect to the scale of 0 to 100. McCroskey (1992) notes that in his conceptualisation, this very broad percentage-type range represents a commonly understood scale. It is difficult to see why a 1 to 10 scale would not be yet more commonly understood.. Of course, in principle the larger scale allows for more gradations of student response. On the other hand, Lingnan students felt initially at a loss which number to choose from, as though the choice had a large element of arbitrariness. Besides, all of them in fact entered scores of 20,30, etc., scores which could just as easily be expressed in single digits. Last, the "Directions" given to the student do not perhaps adequately make it clear that they are being asked to indicate a general willingness to initiate a communication and not to assess their competence so to do. Moreover, many students queried whether they were to imagine they were communicating in English or Cantonese, a factor which would clearly affect their self-reported willingness to communicate.

In short, the instrument worked with moderate success with Hong Kong students and could benefit from modifications so as to ease student confidence. The minor word changes are trivial. The numerical scoring system caused an initial but minor difficulty but can safely be changed to a scale of 1-10. Specifying more clearly that the target language is English would also be useful.

Findings and Discussion

The 12 scored items were entered on a simple spreadsheet and the subscores calculated according to McCroskey & Richmond's (1992) formula.

Table 1

WTC Scores for Pilot Study

Subscores		
Communication Context		
Dyad	44.21	20.61
Group	51.01	20.10
Meeting	45.86	19.18
Public	49.42	20.69
Receiver Type		
Stranger	33.14	20.99
Acquaintance	45.27	22.70
Friend	64.47	21.67
Total WTC Score	47.63	17.93
	Mean	S.D.

These results show certain features we would have predicted. For example, it is common sense that students would be far less willing to communicate with strangers (33.14%) than with friends (64.47%). Somewhat more surprising is the ordering of the communication contexts and the closeness of the mean scores. One would have predicted that students would have been more generally willing to talk one-on-one (dyad) than in a public speaking situation. In this very preliminary study, this seems not to be the case and may suggest that intimate or pair-work contexts are problematic, perhaps for cultural reasons.

There also seems to be a relatively high willingness to communicate in formal public speaking contexts. This may be a feature of the educational expectations put on Lingnan students during their first year where presentation skills are emphasised. The equally high willingness to communicate in groups may also indicate the dominant method of organising student-centred work in the College.

The figures also showed a very high rate of correlation between all the subscores and the total WTC score. This is important since though the individual results for each subscore would be predicted to be different, for the instrument to show content validity, strong correlations would be expected to demonstrate a unidimensional scale (McCroskey & Baer, 1985).

Table 2**Correlation Figures Across Subscores and Total WTC Score**

	<i>Dyad</i>	<i>Group</i>	<i>Meeting</i>	<i>Public</i>	<i>Stranger</i>	<i>Acquaint</i>	<i>Friend</i>	<i>Total WTC</i>
Dyad	1.00							
Group	0.73	1.00						
Meeting	0.75	0.70	1.00					
Public	0.65	0.76	0.75	1.00				
Stranger	0.64	0.72	0.68	0.79	1.00			
Acquaintance	0.82	0.74	0.78	0.72	0.54	1.00		
Friend	0.71	0.76	0.74	0.69	0.44	0.56	1.00	
Total WTC	0.88	0.90	0.89	0.89	0.79	0.86	0.81	1.00

It is interesting to compare the mean score WTC figures with others from around the world using the same instrument and methodology. Of course, there is no expectation that the figures will be similar or consistent since we might anticipate wide cultural and other differences to account for varying results. Additionally, there are as yet no established norms and therefore the results presented below are preliminary.

Table 3**Comparison of WTC Scores Around the World**

	Hong Kong	U.S.A	Micro-nesia	Australia	Sweden	Finland
Total	47.63	65.2	47.3	56.6	58.1	54.6
Dyad	44.21	76.2	49.6	63.8	63.3	72.9
Group	51.01	70.8	55.2	63.3	63.3	59.8
Meeting	45.86	59.7	37.4	53.1	52.2	49.4
Public	49.42	54.2	47.0	46.0	53.3	51.8
Stranger	33.14	38.5	22.9	38.8	37.4	35.1
Acquaint	45.27	72.5	44.4	61.0	62.8	60.7
Friend	64.47	84.7	74.5	75.9	73.8	68.1

This international comparison shows that there is some variety in scoring rates between different countries, though Hong Kong and Micronesia seem consistently low in the overall mean scores. However, there are some real difficulties in making cross-cultural comparisons. The cases of the USA and Australia are obviously different since English is not (for the great majority) a second language. Similarly, comparison with Sweden and Finland may be difficult since the linguistic similarities and cultural closeness between Swedish

and Finnish and English are obviously greater than between English and Asian languages. Cross-cultural comparisons using instruments designed for one particular culture are notoriously problematic.

Nonetheless, taking Table 3 at face value we note that in the various subscores, the context categories reveal that Hong Kong and Micronesian students are the odd ones out in that communicating in groups seems slightly more acceptable than dyadic communication, whereas for the other nationalities, dyadic communication scores highest. This peculiarity might represent a defining characteristic of communication willingness among certain Asian and Pacific Basin communities. Though much more work is required, a hypothesis might be drawn that Asian students, and their cultures in general, tend to low willingness to communicate scores and that one-on-one communications pose difficulties not encountered in other cultures.

However, as already noted, caution should be exercised in making such broad generalisations about Asian students on the strength of inadequate basic data. Klopff (1984), for example, shows that whereas the Japanese had very high levels of communication apprehension, (a construct related to reticence), the Koreans expressed exactly the opposite trend and students from the PRC differed only slightly from Americans.

As a final word on this international comparison, it would appear that Hong Kong students in our sample of 124 Lingnan students evince a relatively low level of willingness to communicate when compared to internationally available data. Whether this a broad trait observable in Cantonese as well is something we don't as yet know. On the other hand, the level of reticence may be linked to the proficiency level of English. Since the proficiency level of English in Hong Kong is regarded by many as lower than that in, say, Sweden, reticence may be linguistically correlated rather than psychologically. These are questions that we do not have answers to. But we may be too far ahead of ourselves. As yet, we do not know if the mean scores represented above are typical of Hong Kong students and until we are confident of establishing a preliminary norm, it would be unwise to speculate.

Next, we look at uncovering any relationship in our pilot group between WTC scores and general performance in English as measured by marks given for the year's work in English. All 124 students took the same English course with a consistent workload, assessment system and common examinations and tests. We would predict a very low level of correlation between general test results in a general English course and WTC total scores. The correlation figures are given below.

Table 4

Correlation Figures Between Total WTC Scores and Course Results

Correlation R	0.168
Regression R ²	0.028
Significant F	0.136

These figures bear out our prediction that there would be no significant relation between our two variables. The r figure of 0.168 is hardly significant and there is certainly no predictive power in the association. That is to say, we could not predict from a total WTC score general performance in English and vice versa. If there is a significant relationship between willingness to communicate and general English performance, it is not borne out using these data.

Conclusion

In this pilot study we have seen that the WTC instrument is broadly acceptable in the Hong Kong context and that with minor modifications it is easy to use. It is certainly worthwhile proceeding with the instrument in order to arrive at a Hong Kong norm of willingness to communicate. At the same time, it might be remarked that the *approach/avoidance* construct is only one of three available to us and that in the future we might wish to score *communication apprehension* rates, or communication anxiety, as an added and important measure of student oral behaviour.

As far as this study goes, we now have an initial set of mean scores and subscores which indicate the relatively low level of willingness to communicate among Lingnan students. This can be used as a yardstick for future experiments. Over a period of time, and with the help of other tertiary institutions in Hong Kong, a picture may emerge of a reliable Hong Kong willingness to communicate norm. A result worthy of further attention is the subscores recorded for dyadic, group, meeting and public communication contexts. Of interest here is whether certain defining characteristics may be evidenced which distinguish Hong Kong students from those of other cultures.

In our results to date, we have arrived at an average mean score of 47.63% on the WTC scale for our sample of 124 students. If we identify those students who scored one whole standard deviation below the mean score, we discover that 23 students fall well below the average and might represent students whose reticence has consequences in classroom activities and in more formal oral test

situations. Though we are as yet far from proving that reticent students will necessarily fare worse than more willing to communicate students in oral situations, there are perhaps grounds for wishing to see all students in a group expressing themselves. If we have the timetabling luxury of placing students in groups of our choosing, we have the option of herding low WTC students together, dotting them around in other groups, or creating extra confidence boosting lessons for them. In this way, our 23 students will be identified as likely to communicate less (and probably less well) than their peers and might as a result be given fewer opportunities to practice language communicative skills. Once identified, the next difficult question is to decide what strategy to adopt in trying to enable them to become more communicative.

References

- Chesebro, J.W., McCroskey, J.C., Atwater, D.F., et al. (1992). Communication Apprehension and Self-perceived Communication At-Risk Students. *Communication Education*, 41, 345-360
- Clement, R. (1987). Second Language Learning and Acculturation: An Investigation of the Effects of Language Status and Individual Characteristics. *Journal of Language and Social Psychology*, 5, 271-290.
- Ely, C.M. (1986). An Analysis of Discomfort, Risktaking, Sociability, and Motivation in the L2 Classroom. *Language Learning*, 36, 1-25.
- Ericson, P.M., & Gardner, J.W. (1992). Two Longitudinal Studies of Communication Apprehension and Its Effects on College Students' Success. *Communication Quarterly*, 40:2, 127-137.
- Kelly, L., & Keaten, J. (1992). A Test of the Effectiveness of the Reticence Programme at the Pennsylvania State University. *Communication Education*, 41, 361-372.
- Klopf, D.W. (1984). Cross-cultural apprehension research. A summary of Pacific basin studies. *Communication Quarterly*, 39, 130-143.
- MacIntyre, P.D., & Gardner, R.C. (1989). Anxiety and Second-Language Learning: Towards a Theoretical Clarification. *Language Learning*, 39, 251-275.
- MacIntyre, P.D., & Gardner, R.C. (1991). Language Anxiety: Its Relationship to Other Anxieties and to Processing in Native and Second Languages. *Language Learning*, 41:4, 513-534.

- MacIntyre, P.D., & Gardner, R.C. (1991a). Investigating Language Class Anxiety Using the Focused Essay Technique. *The Modern Language Journal*, 75, 296-313.
- McCroskey, J.C., Fayer, J., & Richmond, V.P. (1985). Don't Speak to me in English: Communication Apprehension in Puerto Rico. *Communication Quarterly*, 33, 185-192.
- McCroskey, J.C., & Richmond, V.P. (1987). Willingness to Communicate. In J.C. McCroskey and J.A. Daly, *Personality and Interpersonal Communication*, Newbury Park, CA: Sage, 129-156.
- McCroskey, J.C., & Richmond, V.P. (1990). Willingness to Communicate: Differing Cultural Perspectives. *Southern Communication Journal*, 56, 72-77.
- McCroskey, J.C., & Richmond, V.P. (1992). Reliability and Validity of the Willingness to Communicate Scale. *Communication Quarterly*, 40:1, 16-25.
- Young, D. (1986). The Relationship Between Anxiety and Foreign Language Oral Proficiency Ratings. *Foreign Language Annals*, 19, 439-445.

10

IMPLEMENTING QUALITY IN ESP TRAINING: EXPERIENCE FROM THE HONG KONG VOCATIONAL ENGLISH PROGRAMME

Grahame T Bilbow
Hong Kong Institute of Education

Introduction

Most people who are familiar with the state of English language training in Hong Kong know that there is a need to do more to help employees to develop the English they need in the workplace. Visitors to Hong Kong are often struck by the inability of those they meet in service encounters to understand them or to make themselves understood; many employers complain that it is increasingly difficult to get 'good' staff, and are crucially aware of the weaknesses their existing staff have in conducting their business in English.

In general, this Programme (the HKVEP) grew out of the perception that there is a need, a *growing* need as the service sector assumes greater importance, to address employees' work-related English language needs.

The Programme has also grown out of a need *to assure the quality of training provision* by course providers. It is now generally recognised that quality assurance mechanisms, such as ISO9000, are at least as applicable to service industries (including training) as they are to manufacturing industries.

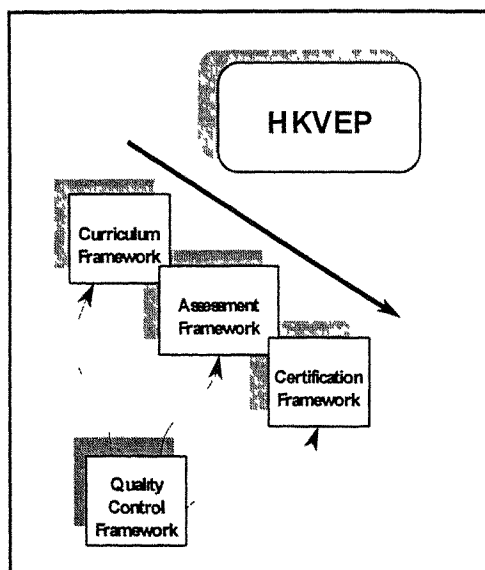
Lastly, the Programme grew out of the realisation that vocational training needs to take account of *future*, as well as present needs. It was necessary, therefore, to create a scheme that would help employees develop their language skills *on a life-long learning basis*.

The Hong Kong Vocational English Programme

The HKVEP is a Hong Kong Government initiative in the area of post-school vocationally-oriented English language training launched in collaboration with the London Chamber of Commerce and Industry Examinations Board (LCCIEB), with the cooperation of a number of training institutions throughout the Territory.

Simply stated, the Programme offers mechanisms for *quality-assuring* and *certificating* training courses in work-related English in the private and public

sectors. Quality assurance is achieved by means of a flexible yet rigorous curriculum and assessment framework applied to courses which teach and assess work-related English. This curriculum and assessment framework owes a lot to work being done in the UK under the banner of NVQs, National Vocational Qualifications, and, in Australia, research in the area of competency-based assessment. We hope that this framework will come to be seen as a quality yardstick offering employers and the person in the street a way of making enlightened choices when considering courses in vocationally-oriented English.



The Hong Kong Vocational English Programme (HKVEP)
Figure 1

A Short History of the Programme

The Hong Kong Vocational English Programme has been fully functional since May of this year, when an agreement was signed between the Hong Kong Government and the London Chamber of Commerce & Industry Examinations Board (LCCIEB). At that point the pilot implementation stage of the Programme, which had been running since September last year, came to an end. An HKVEP Unit has been in existence since May 1993, originally in the Institute of Language in Education and now in the new Hong Kong Institute of Education.

We have two participating training termed Programme Centres. These are the British Council and the Centre for Professional and Business English of Hong Kong Polytechnic University. Since its implementation, the Programme has attracted an encouraging number of students, and since May, more than 600 certificates have been awarded within the Programme.

We have also received a number of enquiries from other training centres interested in becoming approved Programme Centres. This year should see, for example, the Caritas organisation entering the Programme and seeking validation of some of its courses at an appropriate Stage of the Programme.

Quality Assurance Mechanisms within the HKVEP

The subject of this paper is the experience of implementing the curriculum and assessment framework which is at the heart of the Programme. I shall focus particularly on the practical issues we have faced in terms of quality assurance and quality control as they relate to course validation and verification of assessor performance.

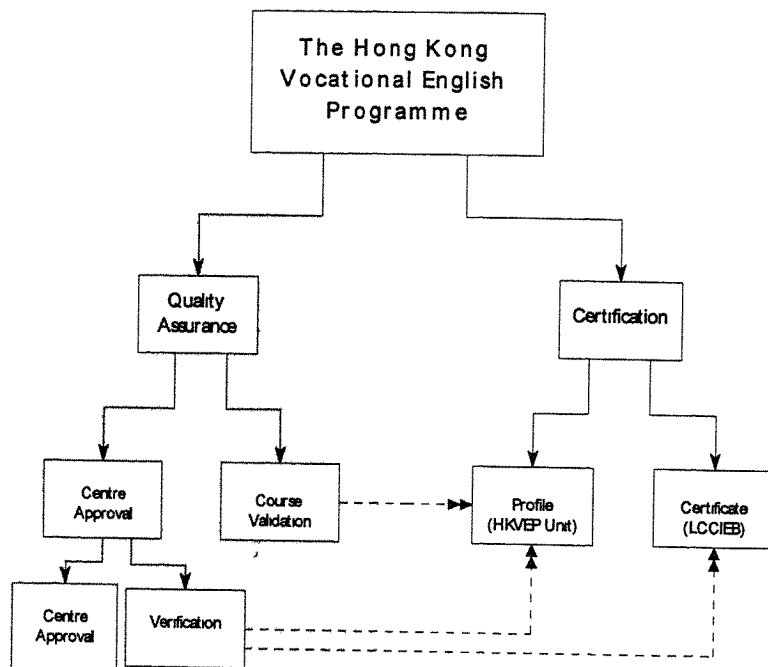
The natural first question to ask is, "what is quality?". Is quality just something that we recognise when we see it, yet which is impossible to define? Although this may in general be true, nevertheless when we talk about quality assurance it is not enough to say, "well, I'll know it when I see it". We have, in fact, to try to define what quality in terms of ESP means in practical terms - in terms of processes and outcomes.

Within the Programme, we define quality in terms of five things These are:

- *Validity*
The aims and objectives of learning are clearly stated, and defined with reference to student needs.
- *Reliability*
Teaching and assessment are reliable and consistent, i.e. on different occasions with different teachers, the results would be likely to be the same.
- *Systematicity*
Teaching and assessment occur against a clear methodological and procedural backdrop. The criteria for determining their success or otherwise are clearly defined.

- *Impartiality*
All students are treated fairly and equally.
- *Value for money*
This is not to be confused with cheapness. Naturally quality costs money. Here, value for money means not exploiting students, creating cartels and monopolies that systematically disadvantage students, and so on.

How do we set about validating courses within the Hong Kong Vocational English Programme against such quality criteria? The following diagram represents the main processes involved in the Programme:



Quality Assurance Mechanisms in the HKVEP
Figure 2

As can be seen in the shaded part of this diagram, quality assurance processes are divided into *centre approval* and *course validation*.

Centre Approval

Although at the moment, the two Programme Centres involved in the Programme are essentially private sector training institutions in the highly-competitive evening-class market, there is no reason why future members of the scheme should not be qualitatively different, e.g. in-house training units of companies, or even schools.

What is an *approved* centre? How do we approve centres? An approved centre is quite simply a centre which we believe is capable of offering validated courses, and of delivering specified awards. In general terms, the purpose of centre approval is to ensure that appropriate quality assurance mechanisms are in place, or can be put in place, insofar as they relate to the teaching, assessment and certification of students.

Centre Approval Visits

During an Approval Visit, a training centre must demonstrate it has, or can develop, the following aspects of quality assurance.

- *Management Systems*
Candidates must have access to adequate information about the training centre itself and the Programme.
- *Administrative Arrangements*
The training centre must demonstrate that adequate systems exist for keeping candidate records and "dossiers of evidence" (*more about this later*). Also a Co-ordinator must be identified whose role is to liaise with the Programme Unit.
- *Physical Resources*
The training centre must have sufficient, appropriate and accessible equipment resources to meet the requirements of teaching and assessing courses within the Programme.
- *Staff Resources*
The training centre must have adequately trained and experienced staff to be able to teach and assess students on courses within the Programme.
- *Assessment*
The training centre must have clear and adequate Assessment Plans, and ensure that adequate information is given to students about the role of assessment within the Programme.

- *Quality Assurance and Quality Control*

There must be co-ordinated internal quality assurance systems. Regular teacher meetings should take place and information should be efficiently distributed among staff. There should also be adequate provision made for Programme Review and Evaluation by students.

The Process of Verification

Centre approval has to be a dynamic process, and should not be restricted to a single centre approval visit. The purpose of the initial centre approval visit is to act as a first step towards a longer on-going relationship between the Programme Unit and the training centre. It is not intended to be threatening, and in some cases, centre approval can be awarded to a training centre which appears to be moving in the "right" direction.

The on-going relationship that I refer to is achieved through *verification* and through collaboration over course design and validation.

Verification is a common term used within the National Vocational Qualification framework and is not quite as sinister as it sounds. Verification visits are made to Centres while validated courses are running, to provide an opportunity for the Verifier to ensure that courses are being conducted effectively and in line with the validated course submission.

Of course, in circumstances where a verification visit reveals that things are not going well, centre approval can be withdrawn, although, wherever possible, it should be possible to negotiate ways out of an impasse.

Verification visits are intended to do two things in addition to the general purpose of allowing Programme Unit staff to maintain contact with Programme Centres:

- 1 They allow the Verifier to observe teaching and assessment taking place. This serves the purpose of allowing the Verifier to determine whether the validated course is being adhered to consistently.
- 2 They allow the Verifier to systematically examine records of student performance. These records are termed, "dossiers of evidence" (another NVQ term). It is necessary to examine these records so as to ensure that:
 - (a) records are being kept efficiently and completely

- (b) there is consistency and reliability of assessment between assessors and within a particular assessor's work
- (c) there is consistency and reliability of assessment between Centres, and between Stages of the Programme.

Verification is largely responsible for ensuring that the criteria for quality mentioned earlier are met.

Course Validation

Apart from centre approval, including centre approval visits and verification, quality is assured through the course validation process.

When a centre is approved, it submits courses to the Programme Unit for validation. Usually, training centres run several different courses on the basis of their analysis of their students' needs. For example, they may run a course in basic business writing, a course in advanced speaking skills for executives, and so on.

Any or all of these courses can be submitted for validation. Validation basically means that courses can then be said to "lead to" certification at a particular Stage of the Programme. There are four Stages, ranging from Stage I, which is concerned with simple workplace-related language activities such as understanding simple instructions in English, through to Stage IV, which deals with complex workplace-related language activities, such as writing complex sales reports in English.

When courses are submitted for validation, they must contain certain elements:

- *Course Details*
This means the title of the course, the total number of contact hours (which must not be less than 50 hours), and the anticipated length of classes (for example two, 1.5 hour classes or one, 3 hour class per week).
- *A Statement of Aims and Objectives, containing:*

Information about the Target Group, i.e. who the course has been designed for, for example, evening-class business English students/ trade apprentices/ junior secretarial staff etc.

A clear indication of students' entry proficiency and the means by which students will be placed in the class, e.g. by placement test, interview and so on.

A Statement of General Course Aims and Objectives, i.e. what the course is aiming to achieve (for example, developing confidence in work-related speaking, developing competence in work-related writing and so on).

Specific objectives in terms of the specific language activities students are expected to have demonstrated performance in by the end of the course.

A Scheme of Work, containing details of:

Course Content & Timing

Information regarding the timing and content of pedagogical focuses i.e. the topics, text-types, skills, non-assessed language activities and language focuses contained in the course.

An indication of the content of assessed activities and the amount of time given to each.

Information about homework content and timing (if appropriate)

Teaching Approach

An indication of the overall approach used to achieve the aims and objectives of the course. In principle, the approach should be communicative, but with focuses on topics, skills, texts and language form as required. We also stipulate that learning should be learner-centred and should encourage learners to take responsibility for their own learning.

Teaching & Learning Resources

Teaching and learning materials used on the course that has been submitted.

A list of teachers responsible for conducting the course.

Assessment

HKVEP assessment is criterion-referenced and based on language produced in carrying out work-related assessed language activities. Students are judged against clearly stated criteria.

A detailed Assessment Plan must be included showing how and when judgments are to be reached. This includes details of the assessed language

activities, including information about the sequencing of assessments through the course.

Conclusion

These two elements of quality assurance, centre approval and course validation are not quite as discrete and "all or nothing" as has perhaps been suggested by the above.

It was mentioned earlier that the purpose of the Programme is to forge links with training centres and to help these centres deliver a high quality product to students of work-related English language.

This, by definition, means that centre approval and course validation tend to be *negotiated* rather than *black and white* processes. This is, after all, as it should be.

On centre approval visits, for example, management and administrative systems are scrutinised. However, at least in the early stages of cooperation, it is quite likely that such systems would not yet exist. This, of course, would not mean that the centre would "fail" the approval process.

Similarly, it is possible that the professional skills and qualifications of staff in an applying institution might fall some way short of the ideal. This would not mean that such staff cannot teach and assess on validated courses. Our aim is to upgrade quality (including the quality of teaching) through negotiation and close cooperation with training centres.

We believe that this framework, together with an effective teacher education programme and on-going support for course providers, will provide a powerful means through which to improve Hong Kong's Vocational English provision.

Looking into the future a little, if the Programme proves successful, it might be extended to other languages e.g. a Vocational Chinese Programme focusing on Putonghua and Written Chinese. It might also act as a model framework for wider vocational education, providing a common but flexible means for certificating vocational education and training in a whole range of subjects across the Territory. If it is to do this of course it will need to liaise closely with institutions currently offering such training, such as the Vocational Training Council.

It is still quite early days for the HKVEP, but we hope that the Programme will be taken up by course providers in other tertiary institutions, in in-company training units, in Government training centres and also in the Vocational Training Council.

11

ASSESSING CORPORATE FOREIGN LANGUAGE NEEDS: CREATING A LANGUAGE AUDITING TOOLKIT ADEQUATE TO MEET INTERNATIONAL AND CROSS-CULTURAL NEEDS.

Nigel Reeves

University of Aston in Birmingham

Language Auditing

A language audit serves two central purposes: to identify an organisation's current Foreign Language Capability, that is

- to draw up an inventory of the languages available
- to what level
- in what posts

and *at the same time* to analyse the organisation's requirement for efficient Foreign Language Communication, then matching that requirement against the assessed capability. The audit can inform both training and recruitment policy. This approach to foreign language training and foreign language capability places them in the same category of importance as conventional corporate planning and management development issues. In the UK, certainly, this has rarely been the case and foreign languages have been treated as a Cinderella among the qualities sought in recruitment or considered in the planning process. The creation of the Single European Market within the European Union has given a new impetus to foreign language learning in the UK, but it has to be said that the main thrust behind the call for the development of sophisticated auditing techniques has come from the European Commission through its LINGUA programme. English still remains the dominant language of European and International trade, though cross-border acquisitions and mergers, together with intensified trading within the mainland of Europe have resulted in the need for more employees to become versed in a whole range of European languages. Competition demands ever greater efficiency in enterprises worldwide, and efficiency in communication is gaining in recognition as a key element. But efficiency and reorganisation techniques such as lean production simultaneously result in calls for targeted training and budgets are certainly not available for massive programmes of general language learning. This explains the interest over the last three years in the development of auditing techniques which make it possible to identify needs accurately

within the context of a company's strategic objectives and, furthermore, to establish closely what training individuals require in order to achieve their goals. It will be apparent that we are talking of LSP not LGP, and that this approach is as appropriate outside the European context as within it. For example, the set of tools which I shall now outline to you would, in my view, be capable of establishing within a Hong Kong enterprise a fair view of future needs for Putōnghua and for Japanese in relation to English. They could further assist enterprises to decide on their training strategy within the Hong Kong Vocational English Programme. Similarly, they could enable an educational organisation to draw up needs specific language curricula for vocational students. In 1991, I had the honour to present a paper in this same series of International Conferences on achieving quality in language learning and training, in which I looked at the application of Total Quality Management approaches to human resource management and to language needs analysis in the education sector. My case study will demonstrate how auditing tools subsequently developed along these lines can be used for vocational language curriculum design and LSP materials production.

The project within which these techniques have been developed is the Business Language WorkBench Project, funded by the European Union LINGUA programme, coordinated by Aston University with the following partners: Mast Learning Systems of London, Coventry Technical College, Faculteit Educatieve Opleidingen HR+O of Rotterdam University, the Netherlands, TEI Kozani, the Technical Education Institution of West Macedonia, Greece, and Escola Superior de Castelo Branco, Portugal.

The Structure of the Business Language Analysis System

As I have mentioned, language auditing as conceived in this project treats Foreign Language Capability as a factor in the corporate planning process, and one of the innovations of the ToolKit is to provide organisations with methods for language needs analysis at a strategic level. The tools are arranged at three levels in all:

- the Strategic or planning level
- the Operational level, which identifies needs at the departmental and functional level
- the Individual level, which is concerned with the requirements of posts and the proficiency of the postholders

The System's ToolKit enables senior management and human resource managers to examine needs from the short to the long term, and to plan training and recruitment accordingly.

Our original intention was to produce an entirely computerised version of the ToolKit, which would not only facilitate the processing of results in large companies but also provide management with a comprehensive expert system to be used as a method for internal assessment of needs. It became apparent in the course of research, however, that the qualitative nature of strategic analysis, together with the very large numbers of variables and permutations of these variables, defied any sensible use of the computer. We have, therefore, opted to offer paper-based or face-to-face analysis techniques where these will provide the most helpful insights most rapidly. Computer-based techniques are used where these are most effective, that is either in areas where quantitative analysis is required, or where the computer software can be integrated within limited parameters. This has proved to be the case for the identification of the defined processes and activities within the organisation's operations that include foreign language communication, and also for recording individuals' perceptions of language proficiency matched against clearly defined proficiency descriptors.

The ToolKit can be entered at any of the three levels, according to the requirements of the user, and a *User Guide* is provided to help with the identification of which tool or tools is the most appropriate. It is still possible for the non-linguist to use the ToolKit, not only to identify needs but, equally importantly, to monitor the validity of training programmes offered by external agencies. By the same token, the ToolKit can be used by specialist linguists to draw up comprehensive, detailed and needs-specific training programmes for industrial enterprises or vocational education establishments.

The Entry Points to the Business Language Analysis System

Briefly, the tools at the Corporate/Strategic Level are intended to identify foreign language capability matched against the global needs of the organisation and its future plans with the consequent language training requirements in mind. Since language training has long-term implications, and has to precede by a considerable period of time the activity that it is intended to facilitate, auditing at this level assumes first importance. The kinds of issue that are addressed in this part of the audit are:

1. Which languages are needed?
In what priority?
In what time sequence?
2. In what business contexts and situations?
3. For what purposes, objectives and goals are the foreign languages used?
4. How will existing and future plans:
 - affect current capability?
 - point to new needs?
 - on what timescale?
5. Are any current problems in foreign markets related to inappropriate language capability?
6. What are the strengths and weaknesses of the enterprise in language resource provision compared with its objectives?

The Operational/Departmental Level tools are designed to identify the departments and functions in the organisation that require or would benefit from a foreign language capability or a higher level of proficiency for maximum efficient working.

They are designed to establish which activities are currently carried out or would most effectively and efficiently be carried out in the foreign language, and which specific language skills are needed in departments and by postholders

Thirdly, they show how often these skills are required and indicate their relative importance, information that is needed for human resource planning.

At the Individual Level, the audit is concerned with analysing individual postholder's language skills and competencies and consequently producing profiles of their language skills, the manner in which they use languages and in what lexical fields.

By means of triangulation, postholders' own perceptions of their proficiency (matched against sets of descriptors) can be compared with the perceptions and experience of their line managers, specialist language auditors or suitably qualified training officers, such as the assessors in the UK National Vocational Qualifications scheme, or the consultants in the Hong Kong Vocational English Programme. The Business Language Analysis System tools allow all these perceptions to be matched against agreed, objective language standards, which are available in both the schemes just mentioned. Furthermore, the posts can be separately calibrated against those standards, so that the training requirement, if any, can be identified.

Returning to an overview of the auditing system, it can be seen that there is a variety of tools for each level at entry-point, and that, for the reasons already stated, they are presented in three categories:

1. computerised
2. face-to-face
3. in a combination of computerised and face to face

Additionally, the RoadMap Tools, and the Language Activities and Documentation Checklists, are available in either form, to be deployed according to circumstances and convenience, the three latter forming part of an integrated tool in the computer-based form.

The philosophy behind the ToolKit is, then, to provide a fully comprehensive system which can be entered at any level, according to need and situation, and with a variety of analytical tools that can be selected in the light of circumstances. To help the user choose the appropriate entry level and to select the most convenient tool, best fitted to the purpose, a *User Guide* has been devised.

The system provides, then, a flexible framework that can be used by large or small organisations to provide broad, long-term information for strategic planning down to detailed information about individuals' immediate training needs. Moreover, it can be used by non-linguists within organisations to evaluate language training programmes offered by outside agencies and provide defined goals for those programmes as a means for monitoring their effectiveness.

I would like to conclude by introducing a case study that shows how the system has been used. The example chosen is taken from the vocational/professional training sector to illustrate how auditing can support needs-specific language curriculum design in the vocational/professional context, and this connects with my 1991 contribution to the QUILT and QUILC conference here in Hong Kong, to which I have already referred.

A Case Study (Vocational and Professional Training Sector)

Background to the Audit

Taining institution in the UK sends students, who are taking a theoretical and practical training programme in the Fine Arts, to Spain for a 3-6 month period. The training occurs partly in lecture rooms, but is largely concentrated in the studio. Students are guided there by Spanish-speaking tutors. Additionally,

they are encouraged to seek possibilities for exhibiting their own works in the city. Spanish students come to the UK under the same conditions.

Purpose of the Audit

The purpose of the audit was to lay the basis for needs-specific training materials and courses in Spanish and English. We therefore had to establish:

how language was used in the learning process
in which vocational situations it was used
what technical terminology was used

Strategic Analysis (Swot Analysis)

In an initial, informal discussion with the institution's senior management, the SWOT approach was used. This was preferred to the Strategic Language Needs Analysis Tool and the Strategic Planning Analysis Tool because the institution had already decided that its collaborative programmes with Spain were a special feature and that long-term importance was to be attached to them. It was, therefore, important to see what impediments there might be to the linguistic level to sustaining and improving this programme as a quality product. The SWOT tool was seen as a rapidly deployable tool for this purpose. It was concluded:

a. STRENGTHS

The strengths of the programme lay precisely in the opportunity for the trainees to extend their learning experience through the introduction to another European Art School, and to extend their artistic techniques and expressive powers through advice from and communication with foreign tutors.

b. WEAKNESSES

The weakness that was recognised prior to the discussion was the inadequacy of the current general purposes language course, and the absence of any needs-specific training materials in the area of the Fine Arts.

c. OPPORTUNITY

The opportunity lay in enhancing the effectiveness of the programme by creating appropriate training materials or techniques.

d. THREATS

The threat lay in loss of prestige through the overseas section of the programme not meeting the high expectations of the trainee customers.

It was concluded from the SWOT analysis that an audit was certainly required to establish precisely what language was used, when, and in what manner during the learning and training process.

Operational Analysis I (Roadmap Tool " Paper-Based)

The operational level audit was carried out on the premises of the institution. The first tool to be used was the *RoadMap*. This had to be modified, since the suggested RoadMaps had been devised with industry in mind. The RoadMap approach was adapted in order to explore the intimate interrelationship between linguistic communication and the vocational/professional learning process. (This was particularly interesting since the products were themselves the outcome of the process, and also a medium of communication; the trainees have to learn to use metalanguage to activate their thinking about the artefact and to interacting others' responses to it). Process analysis therefore seemed essential.

In discussion with tutors, the key stages of the learning and training process were established. There were found to be three concurrent processes, which were interrelated.

1. **A theoretical learning process**, achieved from lectures (listening), from reading and tested through writing. The focus was on the History of Art. This provided the main source of the TERMINOLOGY used for talking about art (*the language of art criticism*).
2. **A practical learning strand**, in which the trainees created their art products in the studio, discussed their papers in dialogue with the tutor (speaking and listening) and, finally, had their art product assessed in a formal *critique* session with an internal or external tutor or professional art critic.
3. **A quasi-commercial learning strand which involved contact with the outside world** at the commencement of the practical process through the purchase or identification of suitable materials, and at the close in seeking an outlet for an exhibition leading to the sale of the art product. This involved negotiations with art galleries and, in some cases, directly with end customers.

Operational Analysis II (Jobholder's Needs Analysis Grid)

Armed with knowledge of the learning process, through the creation of a new RoadMap, it was decided to identify the precise points at which the foreign language was needed, through the use of the Jobholder's Needs Analysis Grid. This was chosen to help the trainees, in a structured group interview, to reflect on and formulate their use of and need for the (foreign) language in their learning and artistic production.

A representative group of trainees, who had already experienced the foreign programme, came together with the auditor and the current language trainer in a brainstorming session.

Since the existing Language Activities Checklist (which is based on language activities in export marketing and transnational manufacturing and trading enterprises) was not appropriate, it was first necessary to establish those activities. Using the outline of the RoadMap, the participants identified the SITUATIONS in which they used the foreign language. These coincided with (and therefore confirmed) the stages of the RoadMap.

Each of these situations was discussed in turn, to establish which language skills were needed — LISTENING, SPEAKING, READING or WRITING (or combinations) — and the specific language activity also identified (eg speaking and listening to the owner of an art gallery to arrange an exhibition; listening and speaking to a tutor to explain an art product; understanding and responding to a critique of an art product; listening to lectures; taking notes; writing short essays.

This newly created Language Activities List provided fundamental information for the Jobholder's Needs Analysis GRID. The trainees then informed the auditor of:

the type of terminology needed in each of these activities e.g.:

the names of raw materials

the expressions of art criticism

the language of selling

thus completing the Jobholder's Needs Analysis GRID.

It was then confirmed which activities required which terminology.

In the event, then, the need to create a new Language Activities Checklist by itself served the function of completing the Activities section of the Jobholder's Needs Analysis GRID.

The trainees were asked to fill in their personal GRIDS after the session for checking. The auditors had completed a "Master" GRID for themselves during the brainstorming session, against which they would later cross-check the individual GRIDS.

This analysis, together with the RoadMap, has revealed the needs-specific language training components for the new training programme. The entire session lasted for 1 hour and 15 minutes. There were 12 trainee participants and 2 auditors. It would not, however, seem any more difficult for the GRID to

be produced by one auditor alone, since the process involving knowledge and use of the foreign language had already been established in the RoadMap analysis.

Work Shadowing

As the auditors were not familiar with the learning processes involved in art studies work, they then spent some 90 minutes attending a series of different studio sessions involving both a range of fine arts operational areas — eg. painting, sculpting and print-making — and a range of learning situations — eg. induction talk, tutor-trainee dialogue and tutor critique of work.

The range of fine art areas provided an impression of the terminology range required.

Observation of the learning situations gave insight into **both** the art production techniques themselves **and** the use of language in the learning/training process.

Conclusion of the Audit

The learning materials now to be created will be based on authentic, video-recorded sequences of the key learning situations/activities identified through the ROADMAP tool. These sequences, together with selected textbooks covering the specialist areas, will provide the basis for an LSP corpus. Since the material is very extensive, the Aston Terminology Analyser, a computerised concordancing tool will be used to assemble the terminology. This enables us to identify the terminology in its typical collocations and to represent them in exercises to facilitate learning. The material will, of course, be checked for terminological appropriacy with the vocational/professional tutors. The same approach will be used in any vocational setting where there is a large, relevant literature. It can, however, also be developed "manually" in conjunction with management.

While video is being used for this language programme, since it gives further weight to the role of language in the learning process, it would be possible to make notes during work shadowing and to base the role plays and exercises practising terminology and language structures on that observation.

Conclusions Drawn from the Audit

The **AUDIT** determined the following, in summary:

The value of obtaining insight into the subject area where the foreign language is used.

This enables training to be based efficiently on

- precisely the activities that require the foreign language competence and the specialist terminology used
- practice that accurately represents the role of language in the process of its related transactions.

The RoadMap Tool proved sufficiently flexible to be adapted to a very specialist service industry, including a complex process, in which communication plays an essential part.

The GRID used in brainstorming manner, proved to be effective in confirming the RoadMap process analysis and in identifying the key situations, activities and terminological areas.

The Work Shadowing method gave the auditors a feel in reality of what had been described and analysed. It is often forgotten by customers that the auditors are very likely not to be familiar with the industrial field concerned. First-hand experience of it, through a visit to the office, plant or workshop can be of enormous help in ascertaining how the foreign language is actually used in practice and within its specialist conceptual context.

- The entire audit was carried out in the space of one intensive half day, demonstrating the efficiency of the techniques.
- Subsequent analysis of the findings took an additional two hours.
- The audit proved a good example of one, paper-based route through the Business Language Analysis System ToolKit, from strategic analysis to the mapping of a training programme.
- The paper-based approach was appropriate because the processes involved in the service industry in question were not recorded in the model RoadMap and Language Activities Checklist provided. The Tools themselves proved appropriate and flexible enough to supply the missing information rapidly.
- The auditors could attend the workplace, and it was not, therefore, necessary in this case to use the computer-based approach.
- It was not the intention of the Audit to establish the Proficiency of the individuals concerned, the purpose of the Audit being to help build a training programme first.

Summary

Language auditing is a very flexible method for gaining insight into an organisation's foreign language requirements. The ToolKit that I have introduced is capable of recording current capability, of mapping that onto needs, both in the present and in accordance with future strategic planning requirements, and of providing the basis for tailor-made curricula and materials production. It can be used in industrial and commercial organisations but, as my case study is intended to demonstrate, also in that most complex of service industries, the vocational education sector. Perhaps of most importance in the context of this conference, it can be used for creating a purpose-built, quality-driven language learning curriculum.

12 MOTHER TONGUE EDUCATION IN A MULTILINGUAL MULTICULTURAL SOCIETY: THE MALAYSIAN CHINESE EXPERIENCE IN THE 1990S

*Ng Kwei Kuen,
Language Centre, University of Malaya,*

Mother tongue education in the Chinese context in Malaysia has been and will always be fraught with controversy in view of the circumstances.

The Chinese make up the largest minority group forming more than 30 % of the total population, which also comprises Malays or Bumiputeras¹, Indians, Kadazans, Dayaks, Kenyang, Bidayu, Melanau and a host of indigenous natives, collectively known as the Orang Asli (see Table 1).

Even by themselves, the Chinese are far from homogeneous. They are made up of diverse dialect groups and are further divided into the Chinese-educated, English-educated and Malay-educated. As pointed out by Professor Wang Gungwu,

"when a Chinese minority is as high as 35 per cent there must necessarily be high concentrations of Chinese in cities and towns and the high concentrations are themselves sources of intergroup and in particular intercommunal tensions." (Wang, 1978)

This complex and divergent make-up of the Malaysian population in general, and the Chinese population in particular, provides the backdrop to the understanding of the problems of Chinese education in Malaysia.

¹ A term used for Malays meaning 'Prince of the soil'. Children of Malay mixed marriages are automatically considered as Bumiputera so long as they follow Malay customs and embrace Islam.

Table 1. Population Profile by Race

	Estimated Total	Malay & indigenous groups *	Chinese	Indian	Others
30 Jun 1990	17,763,000	10,972,000	5,260,000	1,437,000	93,000
30 Jun 1991	18,178,000	11,280,000	5,342,000	1,462,000	95,000

Source: Census May 1992 Monthly Statistical Bulletin Dept of Statistics Malaysia

History and Status of Chinese Education

Malaysians of other ethnic origins have often observed that their Chinese counterparts place a high premium on education. To the Chinese-educated Malaysian Chinese, school is the channel through which the more learned forms of Chinese culture and traditions are passed on from one generation to another. To them, the Chinese language is very often the symbol of the group's uniqueness and identity, of its most cherished traditions and cultural heritage.

"No single institution has been more effective in maintaining a sense of China's cultural heritage than have Chinese schools; their curricula and medium of instruction ensured that Chinese cultural values were transmitted to successive generations of young Chinese." (Cushman and Wang, 1988)

Nevertheless, Chinese education in Malaysia has been through a chequered history as

"the Chinese schools of Malaysia have been steeped in politics from the start of their existence. This has been unavoidable since communalism has been the trade mark of Malayan/Malaysian politics from colonial times. Issues involving some aspects or other of Chinese education have featured in almost every general election, and sometimes by-elections in the Peninsula." (Kua, 1990)

Historically, the British colonial government in Malaya, as the country was known before independence in 1957, was never fully supportive of Chinese education. Policy towards Chinese vernacular education was generally

restrictive. Notwithstanding, by the 1940s, the early Chinese education system from the primary to the secondary level had been established with Mandarin as the medium of instruction replacing the dialects that were used in the beginning dating as far back as 1815.

The post-war period in Malaya saw rapid political changes as the colonial Government began the process of decolonization. Among the changes that took place was a revamping of education policy in relation to language education and culture. According to the 1951 Barnes Report, a new social unity was possible only if the Chinese and Indians accepted that their languages and schools could have no place in the future national system.

The Chinese and Indian communities were told that vernacular education could not be included within the national education system since these two languages were not the official languages of the country. In 1956, the Razak Report recommended that the two public examinations, which marked the completion of lower and upper secondary education, be conducted in only English. This move paved the way for the eventual displacement of Chinese as a medium of instruction in the secondary schools.

However a concession was made for primary schools.

"In 1957, primary schools teaching in Chinese were accepted into the national system together with those teaching in Malay, Tamil and English on the basis that a common curriculum, and not a common language, was the most important basis for national integration." (Tan, 1988)

In 1961 a new Education Act was passed after a review of education policy, which required all secondary schools within the national system to use either English or Malay, the two official languages, as the language of instruction. As stated in the new Act, the ultimate objective is

"for the sake of national unity ... to eliminate communal secondary schools from the national system of assisted schools and to ensure that pupils of all races shall attend national and national-type secondary schools."

In short, under the Government's national education policy, Government-assisted Chinese education was only allowed up to primary level. To enforce the implementation of the new policy, schools which did not fall in line with these requirements had to find their own means of subsistence outside the national system as private schools. The Chinese schools were divided over this;

whether to be co-opted and become part of the state-controlled school system thereby giving up the use of their mother tongue in the teaching of all subjects except Chinese language or to maintain mother tongue education and be financially self-supporting. When the final decisions were made, fifty-four schools joined the national system while sixteen others chose to reject the Government's terms and became independent schools, referred to as "Du Zhong" in Chinese, meaning Independent Chinese Schools (ICS). The Malaysian Chinese schools system has since then been divided into the Government-assisted primary schools and national-type secondary school on the one hand and the Independent Chinese Secondary Schools on the other, thus giving rise to a dual system of Chinese education in the country as shown in Table 2 below.

Table 2. The Dual System of Chinese Schools in Malaysia

Level	Government-Assisted Schools	Private Independent Schools
Primary	Standard 1 to Standard 6. All subjects taught in Chinese with Malay and English as language subjects	Nil
Secondary	National-type Chinese Schools Remove Class* to Form 5 All subjects taught in Malay with Chinese and English as language subjects	Independent Chinese Schools Junior Middle 1 to 3. All subjects taught in Chinese with Malay and English as language subjects Senior Middle 1 to 3. Most subjects taught in Chinese with selective subjects e.g. science and maths taught in English; Malay and English taught as language subjects.
Sixth Form	Lower and Upper Form 6. All subjects taught in Malay with Chinese elective taught in Chinese	Nil

* Pupils from the Chinese or Tamil primary schools are required to go through a one year transition period to get adjusted to Malay as the medium of instruction in secondary school.

The Independent Chinese Schools

To date there are sixty Independent Chinese Schools (ICS) in Malaysia. They are the largest group of private secondary schools in the country. The ICS are

entirely self-supporting, surviving mainly through fund raising and generous contributions from the Chinese community and wealthy Chinese-educated businessmen many of whom sit on the school boards. In the past, the majority of ICS students came from working class and lower income families. However, the trend is slowly changing. Now the student population comes from all walks of life. They range from children of Chinese businessmen, middle class professionals to labourers.

The curriculum of the ICS is limited by conditions laid down by the Malaysian Government which requires the teaching of Malay and English subjects among other stipulations. Some ICS use English to teach certain subjects such as science at the Senior Middle level. In order to make themselves competitive, the ICS also make provision for special classes to be held to prepare students who wish to sit for the Form 3 and Form 5 Government examinations. This means that ICS students are given the option to rejoin the mainstream education system either at the upper secondary level or tertiary level if they so choose. Such a flexible approach in part explains the appeal the ICS hold for many Chinese students in addition to the use of Chinese as the medium of instruction. The value of the ICS is also found in

"the belief that ICS, unlike the National schools, maintain the Chinese tradition of emphasising moral and social education as an integral part of schooling. Thus they preserve Chinese cultural values as well as produce better members of society."
(Tan, 1992)

Winds of Change in the 1990s

If the previous decades (the 50s through to the 80s) were trying times testing the resolve and grit of the Chinese educationists and the Chinese-educated masses, the 90s must then be seen as the age of optimism and hope. Never before has the economy been so buoyant and political climate more tolerant and favourable to Chinese education. An indication of this wind of change is reflected in the sharp rise in demand for Chinese primary education, not only among the Chinese but also the non-Chinese, especially the Bumiputera. According to the latest figures from the Ministry of Education, out of the 30,470 non-Chinese pupils in the Chinese primary schools, 82 per cent are Malays among whom are children from middle class and even prominent family background and 90 per cent of Chinese parents have enrolled their children in Chinese primary schools (Sin Chew Jit Poh, 8 July, 1994) The breakdown of the primary school population by race is given in Table 3 below.

Table 3. Primary School Population by Race , 1994

Type of School	Bumiputera	Chinese	Indian	Others	Total
National / Malay	1,841,227	65,328	78,581	84,107	2,069,243
Chinese	25,095	551,622	5,375	1,733	583,825
Tamil	221	14	104,543	20	104,798
Total	1,866,543	619,964	188,599	85,860	2,760,966

Source: Sin Chew Jit Poh (Malaysia) 8 July, 1994

All of a sudden, Chinese education, at least at the primary level, seems to be given the centre stage. Not only are the Chinese-educated Chinese parents but also English-educated ones, as well as an increasing number of Malay parents, registering their children in the Chinese primary schools. Politicians from both the opposition and ruling coalition parties jump on the bandwagon in appearing to promote Chinese education. More and more, the schools are singled out for their outstanding performance in, for instance, mathematics in comparison to the national primary schools. Mandarin is in vogue. The ability to speak Mandarin no longer carries the stigma it once did. English-speaking parents are proud of their children's ability to pick up the language. Even Cabinet Ministers are openly expressing support for Chinese education. For the first time, the M.C.A. (Malaysian Chinese Association), the component Chinese party in the ruling coalition Government, has volunteered to raise funds for the Chinese Independent Schools in contrast to their previous stand of keeping these Chinese schools at arm's length. Perhaps the change in attitude towards Chinese education is most persuasively expressed by a Malay parent, Encik Suffian, who sent all ten of his children to Chinese school. He said,

"I started doing so (sending his children to Chinese school) 14 years ago. At that time, my children were ridiculed by friends and relatives for the fact that they speak Chinese. But now, my relatives send their children to my house in the hope that they (their children) could learn some Chinese (Mandarin) from my children (Sin Chew Jit Poh, 30 June 1994)

Among the English-educated Chinese parents, too, the change in their sentiment towards Chinese education has been drastic. In the past, especially during the

colonial days, they advised their children to concentrate on English and not to waste time learning Chinese. But now, they insist that their children learn Chinese properly.

Wherefore the Change?

The drama that is being played out in the case of Chinese education in the 90s is indicative of the Southeast Asian Chinese' pragmatic and flexible approach to changing circumstances. Once again it has been shown that

"adherence to Chinese culture and language responds to market forces, and to the requirements for business success and labour's survival in a developing capitalist economy." (Lim, 1983)

Both domestically and internationally, rapid changes are taking place. To begin with, the Government under the current leadership has embarked on an accelerating pace of development. The Prime Minister, Dr. Mahathir, started steering the country towards closer economic and trade relations with her Asian neighbours with the "Look East Policy" in the 1980s following which close economic ties were forged with Japan in the hope of benefitting from the transfer of technology from Japan, the most powerful and technologically advanced Asian nation. Subsequent to Japan, the emergence of the "four dragons", Taiwan, Korea, Hong Kong and Singapore as the fastest growing economies in the world also pointed the way of things to come in Malaysia. In the early 90s the grand Vision 2020 was announced by the Prime Minister as the way forward in transforming Malaysia into an industrialised state by the year 2020. In order to ensure success in achieving the objectives set out in the Vision, the Government began to adopt more liberal policies. The success of the NEP (New Economic Policy)² also allowed liberalisation of Government policy to proceed more smoothly as reflected in the figures below.

"In 1979, Bumiputera accounted for only 11% of the architects, 7.6% of accountants, 11.6% of engineers and 8.6% of doctors in both the private and public sectors." (Fourth Malaysia Plan 1981-1985)

² The NEP was introduced following the 1969 racial riots. The NEP provides the Government with the rationale for the restructuring of society. Specifically, Bumiputeras are given preferential treatment in all spheres of development in the Government's attempt to eradicate inequality of opportunity along racial lines.

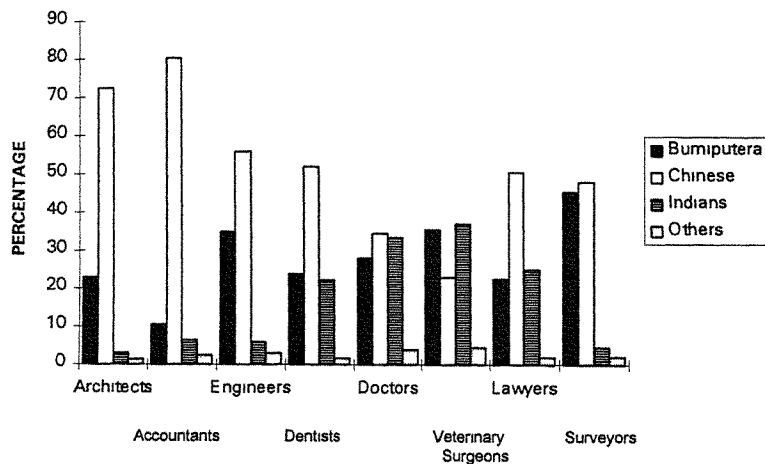
By the Sixth Malaysia Plan 1991-1995,

"of the total 996,400 new jobs generated during the period, the Bumiputera gain as much as 60% compared with Chinese, 32% and Indians, 7.0%. In terms of occupational categories, the majority of Bumiputera workforce was still in the lower-paid job categories. However, their share in certain higher-paying professional occupations such as engineers, accountants and architects increased from 22.2% to 29%." (See Chart below)

In the restructuring of the corporate sector, the share of Bumiputera, including trust agencies, accounted for 20.3% of the total paid-up equity capital in the corporate sector in 1990 as opposed to 9.2% and 12.4% in 1975 and 1981 respectively.(Fourth Malaysia Plan, p.62) The need to make Malaysia competitive together with the rising confidence of the Bumiputera has contributed to the downplay of racial politics and has set the focus and tone of development instead on economic co-operation, domestically, among the various races, especially in Bumiputera-Chinese partnership and internationally, between Malaysia and the world, particularly the East Asian neighbours. In education, this has resulted in an unprecedented openness towards things Chinese, such as Confucian values, and Chinese education..

REGISTERED PROFESSIONALS BY ETHNIC GROUP, 1990

Source. Sixth Malaysia Plan 1991-1995



This liberalisation arising out of the general sense of achievement and confidence on the part of the Malays seems to confirm the perceptive view of Prof. Wang Gungwu who suggested that

"cultural and ethnic differences in themselves are secondary. Far more important are those activities which lead to economic competition, especially when they bring into prominence the values and attitudes which that economic competition loaded in favour of the Chinese against the indigenous peoples.(Wang, 1978)"

In the meanwhile, as the economy of the four Asian "dragons"- Taiwan, Hong Kong, Korea and Singapore grew from strength to strength, the Malaysian Government wasted no time in wooing investors from these countries. Taiwan responded most positively in the late 1980s with rather substantial investments and has since become Malaysia's fourth largest trade partner. Hitherto, the total Taiwanese capital investment in Malaysia amounts to more than US\$60 billion, the highest in Southeast Asia. The economic value of the Chinese language was soon realised as seen in the great demand for Malaysian workers who are proficient in Chinese among the Taiwanese firms and factories. Interestingly enough, the Government in their effort to entice Taiwanese capital into the country cites the language advantage Malaysia has over the other countries in the region, apart from Singapore, as an added attraction to do business in Malaysia. The economic value of the Chinese language has been further enhanced in the last few years with closer ties between Malaysia and the "awakened giant", the People's Republic of China (PRC). The liberalisation of the PRC's economy has opened up a vast potential market. Malaysian entrepreneurs have been quick to seize the opportunities and have made substantial investments in the PRC. Trade missions, state visits by the top leaders from both sides are occurring more regularly. Leaders of both Malaysia and China have conceded that the relationship between Malaysia and China has reached an unprecedented high point. In the last 20 years, trade figures have risen almost twelve fold, jumping from a mere US\$1.5 billion in 1975 to US\$17.7 billion in 1993. (Sin Chew Jit Poh, 30 May, 1994)

Problems, Challenges and Prospects

What does the future hold for Chinese education in Malaysia? In the short to medium term, it obviously looks rosy. However an objective appraisal of the difficulties and obstacles faced by the Chinese schools in particular and the Chinese community in general and how these are inter-linked to the larger socio-economic and political picture should precede any conclusions.

A Real Threat - The Education Act of 1961

Section 21(2) of the Education Act 1961 is by far the most worrying threat to the future of Chinese education in Malaysia. It states:

"Where at any time the Minister is satisfied that a National-Type primary school may suitably be converted into a National primary school he may by order direct that the school shall become a National primary school." (Source: The National Language Bill, 1967, Kuala Lumpur, Government Press, Sections 3 & 4)

The implications of this clause have not escaped the Chinese community. Clearly the National-Type Primary Schools are the nurseries of the Chinese Independent Schools. In the event that Chinese primary schools are abolished, private Chinese education (the ICS in particular) would be dealt a mortal blow. This clause also clearly implies that mother tongue education in Malaysia is not given any constitutional protection and is not considered a democratic right of minorities. As a result, the Chinese schools are put in a very vulnerable position. Repeated appeals for repeal of the clause have thus far not been entertained. The Chinese educationists look upon the Act as the sword of Damocles that hangs above the neck of the Chinese schools. There is no sign of their misgivings being lifted.

The only visible glimmer of hope in the 1990s is that Dongjiaozong³ and the Chinese component political parties (the MCA and Gerakan) of the ruling coalition are moving towards greater co-operation in their effort to secure a legitimate place for Chinese education in the country. This is seen in the improved relations between Dongjiaozong and the MCA who have never seen eye to eye prior to the 1990s, with the former accusing the latter of 'selling out' Chinese interests. Recently, there was a change of leadership in Dongjiaozong. There is speculation that the new leaders may adopt a more pragmatic approach in fighting for the cause of Chinese education. Whether or not the new leadership of Dongjiaozong can succeed in making a breakthrough is left to be seen.

³ Dongjiaozong comprises Dongzong, the United Chinese School Committees Association and Jiaozong, the United Chinese School Teachers' Association. The two bodies have worked closely in articulating and mobilizing Chinese opinion on education and a wide range of other issues.

Differential Treatment of Schools

The Chinese schools view themselves as being given the "step child" treatment in comparison to the national schools in terms of Government support. For instance, the Ministry of Education is not being seen to be taking any concrete action towards addressing the desperate twin problems of shortage of Chinese language teachers and Chinese primary schools in densely populated urban areas. In the last twenty-three years, the number of Chinese primary school pupils has increased by 100,000. However, this considerable increase in student population is not matched by a corresponding increase in the number of schools built. This shortage of teaching staff and school buildings has reached critical point in many schools. When places in the Chinese primary schools are filled up, the pupils who are turned away are placed in the national primary schools much to the resentment of disgruntled parents who are forced to see their children attend schools which they believe to be inferior.

In addition, facilities in many Chinese schools are said to be wanting and schools are in urgent need of repair. However, more often than not, requests for financial assistance fall on deaf ears. It is not without reason that, the Chinese population remains sceptical of the Government's sincerity and view official statements of intent, or verbal support for Chinese education as mere lip service.

Competitive Edge Of Chinese Schools

In the context of the 90s, Malaysians need to be trilingual to be competitive. In this respect, theoretically speaking, the Chinese schools have the edge over the national schools as their students have to study all three major languages, Chinese, Malay and English. Practically, however, the acquisition of language skills is often an uphill task for the majority of students. As a whole, Chinese-educated students are not as proficient in Malay and English as they would like to be, even though the relevance of these two languages is widely recognised. A 1992 survey of 51078 Junior Middle and Senior Middle ICS students with regard to their self-assessment of language proficiency revealed that less than 10 per cent of the students consider themselves capable of doing well in Malay and English (Table 4). Apart from the Chinese language, most are not confident of their linguistic ability in Malay and especially English.

Table 4. Self-Assessment of ICS Junior and Senior Middle Students on their Ability to obtain Credits in the Chinese, Malay and English Language Subjects- Peninsula, Sarawak and Sabah, 1992 in Percentage

Subject	West Malaysia		Sarawak		Sabah	
	Junior Middle	Senior Middle	Junior Middle	Senior Middle	Junior Middle	Senior Middle
Chinese Language	39.49	39.48	42.58	39.77	34.94	35.42
Bahasa Melayu*	5.13	3.04	4.23	2.05	7.63	5.59
English Language	5.43	4.78	5.47	4.70	6.61	5.66
Chinese & B.M.	8.76	4.90	6.39	4.91	10.76	6.72
Chinese & English	7.3	7.72	7.00	4.91	8.19	4.31
English & B.M.	4.57	2.48	3.16	1.88	7.45	5.21
Chinese, English & B.M.	8.97	5.50	6.85	4.61	9.95	5.59
Neither Chinese, English nor B.M.	20.36	32.10	24.33	37.16	14.48	31.50
Total	100	100	100	100	100	100

Source: Dongzong (published in Sin Chew Jit Poh, 29 June, 1994)

Where reading ability is concerned, while the majority of the West Malaysian ICS students can read and write fluently in Chinese, only about a third felt they can do the same in the national language. An even smaller percentage, 24 per cent, felt they can read and write well in English. ICS students from the two East Malaysian states are in more or less the same situation. For instance, about 50 per cent of the Sabah ICS students in Junior Middle and Senior Middle school can read and write fluently in Bahasa Melayu, while 34.87 per cent of the Junior Middle school and 28.92 per cent of the Senior Middle students can perform similar tasks in English. As for speaking the two languages, a great majority find difficulty in expressing themselves orally in the two languages out of fear or inhibition, more so in the case of English than the national language. (Sin Chew Jit Poh, June 29, 1994)

Unless the Chinese schools can upgrade their students' language proficiency in English and Malay substantially, any claim of trilingual advantage will remain unsubstantiated. If the Chinese schools are unable to produce results in terms of the effectiveness of their teaching, no amount of cajoling can attract more students to enrol.

Racial Bias

Owing to the racially delicate situation in Malaysia, the fight for the democratic right to mother-tongue education can be misconstrued as a chauvinistic demand. There is a lot of historical baggage that needs to be overcome on both sides. The tendencies of the Chinese as well as the Malays towards ethnocentricism have to be eliminated. Undeniably, there exist

"differences in thought patterns, values and attitudes towards work and profit, and even attitudes towards education, material advancement and social mobility. And not least, there is the factor of ethnic difference with the various biases and prejudices concerning intelligence, health and moral purity which follow from deep-seated racial prejudice." (Wang, 1970)

It has been pointed out that the Chinese tend to suffer from a sense of cultural superiority vis-à-vis the indigenous peoples on the one hand and feeling inferior to the Westerners on the other (He Guozhong, 1994). If such self-defeating attitudes are left unchecked or worse still, encouraged, they will almost certainly aggravate the mutual distrust and animosity among the races. To adopt a more confident and self-affirming approach to issues concerning Chinese language, Chinese education and Chinese culture, the Chinese-educated community must first and foremost discard their "defensive psychology" so that they will learn to accord more respect for others and to initiate intercommunal interaction. For the most part Malaysian Chinese are slow to integrate with other ethnic communities.

The Danger of Economic Exclusion

Most people seldom study language for its own sake. What matters is the instrumental or economic value and hence control that language wields. The situation in Malaysia has certainly testified to this truism. As long as the existing extensive business network of the Chinese business community remains intact, Chinese language will continue to provide the Chinese-educated with the leverage to economic activities and participation and a secure niche in the predominantly Chinese business sector especially at the lower retail trade level. Conversely, entry into the Chinese business sector would, by default, be denied to all others not equipped with a working knowledge of the Chinese language. If used deliberately as a defensive weapon, this exclusion could be expected to keep the Chinese even farther apart from other races.

As Chinese begins to assume more importance as the regional business language, the other ethnic groups, especially the Malays, are motivated to learning it, as evidenced by the new phenomenon of Malay children attending the Chinese primary schools. While this new development ought to be taken as a much welcome change in the right direction, unfortunately, there are some Chinese who are less than enthusiastic about it, partly out of fear of competition, partly as a reaction to the perceived unjust implementation of the NEP and partly perhaps owing to sheer myopic self-interest. With the Malays and other ethnic groups knowing Chinese, some fear, the Chinese may ultimately be deprived of even their last competitive edge, their language. While it is true that all racial groups act out of self-interest to compete for more socio-economic and political control and benefits, what is of social consequence is how such self-interest manifests itself. It is pertinent to consider if self-interest is better served with each ethnic group pitting itself against another or would it be more worthwhile finding common goal for the realisation of such self-interest. Malaysians and the Malaysian Chinese in particular must learn to realise that sharing is not only desirable but also feasible.

Conclusion

The Chinese school will thrive as long as the Chinese community continues to support it. This support in turn is dependent on the socio-economic mobility Chinese education can offer to the Chinese population in Malaysia. The 1990s seem so far to be the "best of times" for the development of Chinese education in view of the buoyant economy and favourable political climate. However, given the complex multi-racial, multi-cultural nature of the society, the future of Chinese education cannot rest solely upon the subjective resolve of the Chinese community in their fight for a rightful place for Chinese education in the national education system. The challenge the Chinese community in Malaysia must seriously ponder is, as Adam Makkai puts it:

How do Malaysians, the Chinese Malaysians in particular, practise a kind of 'moral fraternity' that would embrace one and all and does not exclude anyone because of his language, faith or skin colour? (Makkai, 1987)

There is no easy answer to the difficult question. The problems in relation to Chinese education before the Malaysian Chinese are as many as they are intractable. Subjective perceptions of ethnic differences and conflicts of interest remain and are heightened by state ethnic policies and ethnic politics. Notwithstanding, what is clear to an objective observer is that "Chinese

Malaysians cannot shape their own identity, culture and socio-economic rights purely by focusing on their own ethnic interests alone." (Tan, 1988)

As much as the Chinese educationists are fighting against hegemonic domination of Government policies on culture and education, they must also guard against the pitfalls of self-centred ethnocentrism and racial arrogance. If armed with critical and objective self-knowledge and a firm grasp of the political reality in the country, the Chinese education proponents may eventually relieve themselves and the community of the historical baggage in connection with the struggle for mother tongue education. Would the Chinese-educated Malaysian Chinese be able to capitalise on the current favourable conditions and work towards a constructive conflict resolution acceptable by all ethnic communities? Only time will tell. What may be said for now though is that time seems to be on the side of the Malaysian Chinese as the world awaits the coming of the 'Age of The Pacific Rim' in the 21st. century. As Asian countries, especially the 'giants' like China and Japan, gain more power and say in economic matters and world affairs, Asian languages can be expected to be more widely used. This trend of development at the international scene should augur well for Chinese education in Malaysia. But in the final analysis the future or prospect of Chinese education in Malaysia is governed more by the development inside rather than outside the country.

References

Cushman, Jennifer and Wang, Gungwu 1988 Literacy and Culture (Introduction). In Jennifer Cushman and Wang Gungwu (eds.) *Changing Identities of the Southeast Asian Chinese Since World War II*. Hong Kong: Hong Kong University Press.

Federation of Malaya 1951a Report of the Committee on Malay Education (Barnes Report)

_____ 1956 Report of the Education Committee (Razak Report)

_____ 1957 Education Ordinance

_____ 1961 Education Act

Fourth Malaysia Plan 1981-1985 1981 Kuala Lumpur: The National Printing Department.

- He Guozhong 1994 Li Xing de Shiluo - Lun Dang Dai Ma Lai Xi Ya Hua Ren de Min Zhu Yi Shi He Wen Hua Ren Tong. Paper presented at the International Conference on the Direction of Chinese Culture for the 21st Century. Kuala Lumpur.
- Kua, Kia Soong 1990 A Protean Saga: The Chinese Schools of Malaysia. Kuala Lumpur: the Resource and Research Centre, Selangor Chinese Assembly Hall.
- Lim, Linda Y.C. 1983 Chinese Economic Activity in Southeast Asia: An Introductory Review. In Linda Y.C. Lim and L.A. Peter Gosling (eds.) *The Chinese in Southeast Asia, Vol. 1, Ethnicity and Economic Activity*. Singapore: Maruzen Asia.
- Makkai, Adam 1987 Cultural Dominance and Rights of Minorities. In Kua Kia Soong (ed.) *Defining Malaysian Culture*. Kuala Lumpur: K. Das. Ink.
- Ng, Kwei Kuen 1983 Language, Ethnicity and Economy among Urban Chinese Squatters- a Preliminary Survey. *Man And Society*, Vol. 4, 1983, Journal of the Department of Anthropology and Sociology, University of Malaya.
- Sin Chew Jit Poh (Malaysia) January 1994 - December 1994.
- Siow, Moli 1983 The Problem of Ethnic Cohesion among the Chinese in Peninsular Malaysia: Intraethnic Divisions and Interethnic Accommodation. In L.A. Peter Gosling and Linda Y.C. Lim (eds.) *The Chinese in Southeast Asia, Vol. 2, Identity, Culture and Politics*. Singapore: Maruzen Asia.
- Sixth Malaysia Plan 1991-1995 1991. Kuala Lumpur: The National Printing Department.
- Tan, Chee Beng 1988 Nation Building and Being Chinese. In Jennifer Cushman and Wang Gungwu (eds.) *Changing Identities of Southeast Asian Chinese Since World War II*. Hong Kong: Hong Kong University Press.
- Tan, Liok Ee 1988 Chinese Independent Schools in West Malaysia: Varying Responses to Changing Demands. In Jennifer Cushman and Wang Gungwu (eds.) *Changing Identities of Southeast Asian Chinese Since World War II*. Hong Kong: Hong Kong University Press.

1992 Dongjiaozong and the Challenge to Cultural Hegemony 1951-1987. In Joel S. Kahn and Francis Loh Kok Wah (eds.) *Fragmented Vision- Culture and Politics in Contemporary Malaysia*. Sydney: Allen & Unwin.

Wang, Gungwu 1978 *The Chinese Minority in Southeast Asia*. Singapore: Chopmen Enterprises.

Wang, Gungwu 1970 Malaysia: Contending Elites. *Current Affairs Bulletin* (University of Sydney), 47, iii.

13

THE QUANTUM LEAP: FROM ESSAY TO THESIS

Linda Cooley, Jo Lewkowicz

English Centre, University of Hong Kong

Until comparatively recently it was assumed that students taken on for a research degree would have the necessary linguistic skills to cope with writing their thesis. Students were expected to make the transition from writing at undergraduate level to writing a postgraduate thesis with little help or guidance on writing. The help that was forthcoming was normally limited to advice from the subject-specialist supervisor who would not necessarily understand the difficulties students were facing if they were not related to the nature of the research per se. However, with a rise in the number of research students together with increased pressure from funding agencies for students to complete their degrees within the allocated time, there is now a growing interest in the reasons why students fail to complete (see, for example, DEET, 1988 and Dunkerley and Weeks, 1994). There is also a growing realisation that thesis writing is a unique experience with which students, even those writing in their L1, may need some help. A research thesis demands a whole new approach to writing, particularly for those students from disciplines where undergraduate essays are generally short and extensive revision is not required. Students trained in a knowledge-telling writing style often have to make the transition to a writing style which allows them to negotiate their knowledge claims in the light of those of others. They have to make the leap from essay to thesis writing. (For a good summary of the demands of thesis writing see Torrance and Thomas, 1994.)

To help students cope with these demands a number of reference texts have appeared on the market (these include: Philips and Pugh, 1987; Hamilton, 1990; van Wagenen, 1991). They cover areas such as 'Writing the Thesis' and 'The Style of Dissertations', which though useful, have been more often than not treated very generally. The books, by trying to address the whole spectrum of students from the Humanities to Medicine, have frequently failed to go beyond raising students' awareness of what is expected of them. They address the *dos* and *don'ts* of thesis writing rather than specific writing problems which students face as they write. They have therefore been seen by many to be of limited use, particularly in situations where they are the only provision available to the students.

Recent increased understanding of the demands of thesis writing has brought a recognition of the need to provide research students with professional help

with their writing. In the US where students take taught courses at the beginning of their PhD. programme, this help frequently comes in the form of taught courses such as the one developed by Swales and Feak at Ann Arbor University and published as *Academic Writing for Graduate Students* (1994). The obvious advantage of a taught course is that it is cost-effective and may be made compulsory for all students. However, a course may fail to address the specific needs of the students as and when they arise, particularly when it is part of a requirement to be fulfilled prior to the students embarking on any extensive writing for their own research.

Alternatives to general writing courses include workshops and seminars based on the writing students are already producing for their supervisors. Such alternatives can address specific issues that students identify as problematic, as can individualised help which, though expensive, may be seen to be of greatest benefit to the students. These are the directions that universities appear to be following in the UK where compulsory taught courses for research students are still uncommon. The drawback of such voluntary programmes is that students may be reluctant to seek advice since they may fear their writing difficulties will be misinterpreted as difficulties in their conceptual understanding of the subject (Torrance and Thomas, 1994: 106). Yet working on writing they are doing for their supervisors is more likely to be seen by students as relevant to their needs.

Since writing difficulties have been recognised as a major issue at universities where a minority of students and supervisors are working in a second or foreign language, it is conceivable that the scale of the problem is greater where the majority are using an L2 or FL. Yet at the University of Hong Kong (HKU) which faces precisely this situation, all the evidence until recently had been anecdotal rather than systematically researched. No provision for research students encountering writing difficulties was made until 1992-93 when the English Centre introduced the Writing Support Service. It was in response to the demand for this one-to-one consultation service that the School of Research Studies requested at the beginning of the 1994-95 academic year a more systematic investigation into issues concerning postgraduate writing.

Background to the HKU Project

To ascertain the extent of the writing difficulties facing research students it was deemed necessary to look at students' perceptions of their own writing problems, and supervisors' perceptions of their students' problems as well as extensive samples of student writing. It was felt that such a three-pronged approach would give the most complete picture of difficulties facing current

students and at the same time would enable the project team to develop an analytic framework which could be used for diagnostic purposes in the future. The number of students registered for research degrees (MPhils and PhDs) at the University is in excess of 1200, and of these the majority (over 90%) are required to submit their thesis in English. Students are registered in one of nine faculties and though some faculties have significantly fewer research students than others (notably law and architecture which are predominantly providing professional training), it was considered important that the views of both students and supervisors across all faculties were sought.

Students' views are being targeted through a questionnaire which has been designed to elicit a self-rating of the students' writing ability in English, the extent of their experience in writing in English and the areas with which they perceive themselves to have difficulties. Information from the questionnaires will be used to determine whether there are common problems among students and if so, whether these are common across disciplines or related to students' past learning experiences in terms of the language of instruction for their first degree. (A significant minority of the research students at HKU are from the PRC and they have not been through an English-medium education). The information will in addition be used as a basis for designing a series of experimental workshops to be run for the students over the next two years.

Student writing samples of at least 10-20 pages have been collected from a wide range of sources. Some are from students who have sought assistance from the English Centre. Others are from students who may consider themselves, or who may be considered by their supervisor to have few writing difficulties. These are being analysed to determine the common strengths and weaknesses of students' writing and they should prove a valuable source of data.

A third source of input to the project is from views on students' writing elicited from supervisors. These views are being gathered through a structured interview which is divided into four parts: personal data; perception of the level of English of current students and their writing difficulties; the linguistic demands of the field in which the supervision is taking place and the writing requirements that students need to meet; and finally supervisors' views on the provisions that the University should make for students to help them with their writing. The interview lasts approximately 30 minutes and is, whenever practicable, recorded.

A structured interview was chosen in preference to a questionnaire to ensure that views were elicited from supervisors across all nine faculties, from those whose L1 is not English as well as those whose L1 is English, and from

supervisors with considerable experience in terms of supervision as well as from those with limited supervisory experience. The interviews had to be structured to ensure consistency among the interviewers. Even so, it is possible that by placing different emphasis on certain questions responses may, to some extent, depend on which member of the team is carrying out the interview. Of the c500 supervisors, it is hoped that 20% will be interviewed and their views will be used to inform the needs analysis. At the time of writing, 10% of the target population have been interviewed and it is to their perceptions of students' needs that the rest of this paper turns.

Analysis of Findings

Since the analysis of the findings presented in this paper is based on interviews with 46 supervisors, that is 50% of our target number, the findings reported here are, naturally, rather tentative and certainly open to modification in the light of later investigations.

(i) Breakdown by Faculty

Although we were striving initially to have a fairly even distribution of supervisors across faculties, Figure 1 shows that this aim was not achieved. Architecture and Dentistry are somewhat under represented and no supervisors from the Law Faculty have yet been interviewed. This reflects the relatively few students registered for postgraduate degrees in these faculties. Indeed, it seems that to have an even distribution across faculties will be impossible, but we will endeavour to include some supervisors from Law in our final analysis and a number from Architecture and Dentistry proportional to the numbers of students supervised in these faculties.

(ii) L1 and L2 speakers of English

A breakdown of supervisors into L1 and L2 speakers of English (Fig.1) shows quite an excess of L1 speakers over L2 speakers. Our intention is to redress this balance as much as possible by targeting L2 speakers in later interviews so that the numbers in the two groups are as close as possible for the purpose of comparison, should any differences become apparent which may seem worthy of investigation at a later stage of the study.

(iii) Language background of students under supervision

The supervisors interviewed so far were responsible for 95 MPhil. students (including 9 L1 speakers of English) and 131 PhD. students (19 L1 speakers). The L1 speakers of English are concentrated in the Arts and Humanities. Only 3 of the 9 L1 MPhil. students are in scientific or technical disciplines (1 in science and 2 in Dentistry) and of the 19 L1 PhD. candidates, 18 are in the Arts

and Humanities. As mentioned earlier, this relatively unusual situation of having such a large majority of students working in an L2 or FL may considerably increase the scale of the problems faced by students when writing a dissertation.

Figure 1: Breakdown of numbers of supervisors by faculties and into L1 and L2 speakers

FACULTY	Total	L1	L2
Education	6	5	1
Architecture	3	2	1
Arts	7	6	1
Dentistry	2	2	0
Social Science	8	6	2
Engineering	5	2	3
Science	8	2	6
Medicine	7	3	4
TOTAL	46	28	18

(iv) Numbers perceiving language problems

Not all the supervisors interviewed considered that their current students have difficulties when writing in English. In fact, 7 of the 46 (15%) did not perceive any difficulties, (14% of L1 speakers and 16% of L2 speakers) leaving 85% who did identify some problems (Figure 2).

It is perhaps interesting to observe that the students of the 7 supervisors who did not perceive any difficulties were not, for the most part, L1 speakers of English. It was not, therefore, the fact that the students were writing in their first language that accounted for the lack of problems. What did seem of significance in some cases, in Architecture and Surveying, for example, where 2 of the 3 supervisors perceived no problems, was that the students were mature students who had been working in the medium of English before taking up research. In Education the students again were mature and were actually doing research into English language or the teaching of English. They were, in fact, the students one would least expect to have problems.

Figure 2: Number of supervisors perceiving problems

FACULTY	Total	Problems	
		YES	NO
Education	6	4	2
Architecture	3	1	2
Arts	7	7	0
Dentistry	2	2	0
Social Science	8	6	2
Engineering	5	5	0
Science	8	7	1
Medicine	7	7	-
TOTAL	46	39	7

(v) Problems Perceived

We now turn to look at the types of difficulties identified by the supervisors. For this purpose, only six of the faculties in the university were included in the full analysis. Architecture and Dentistry being so under-represented in the survey at this stage were excluded. In fact, as already noted (Fig. 2), only 1 of the 3 supervisors in Architecture identified any writing problems and only in 2 areas and both of the supervisors in Dentistry identified problems in 1 area only. We will refer to these as the difficulties are discussed, but it seems inappropriate to include these faculties in the analysis at this point.

The perceived problems have been categorised into seven groups on the grounds of the frequency with which they were mentioned rather than in terms of a hierarchy of severity. Four of these were quite easily categorised as distinct features, but the remaining 3 were a little problematic; the problem being in deciding whether or not they do form distinct groups or are subsets or indeed combinations of the other 4 categories.

We begin by presenting a breakdown of the four distinct categories and then deal with the other 3 to try and explain where the problem lies. The most frequently mentioned problems have been categorised as SURFACE STRUCTURES (Fig.3). This group had a fairly high rating in all faculties, though it appears to be perceived as less of a problem in the Arts disciplines. This was also the one area of writing which the supervisors in the faculty of Dentistry perceived as a problem. Perhaps it should also be mentioned here that

this was the only one of the categories identified by the supervisors where L1 speakers of English were perceived not to have problems.

One reason which may explain the frequency with which supervisors mentioned errors such as misuse of the definite article and incorrect subject/verb agreement is the fact that this is an area of difficulty which supervisors who are, for the most part, non-language specialists can readily recognise and one which they feel confident in dealing with. However, it should be stressed that several of the supervisors did confirm that these surface errors could not be considered a major problem in so far as they rarely obscure meaning. They can cause temporary confusion for the reader and thus make the writing less efficient than it should be, but they are more a source of annoyance than an obstacle to understanding. (See Casanave and Hubbard 1992 for a survey of supervisors' ratings of such features.)

We have yet to find out from looking at extensive samples of student writing across the faculties if surface structure errors are, in fact, as widespread as supervisors seem to believe they are. And also if there are differences between faculties in the frequency of occurrence of errors in surface structures.

It will also be interesting to compare the frequency with which the research students perceive themselves to have difficulties with the type of grammatical problems that the supervisors in our survey noted. A very cursory analysis of our student questionnaires in this study suggests that grammar was not, in general, seen as a problem area. (See Shaw 1991 for students' self-assessment of grammar problems). Perhaps, of course, it is this very lack of perception of the problem that causes it to continue; a problem must be recognised before it can be corrected.

The second group of problems identified has been categorised as those relating to LOGICAL DEVELOPMENT (Fig.3). Quite a wide variety of terminology used by supervisors has been grouped together here ("poor organisation of ideas", "inability to present a logical argument", "incoherence" etc.) as they all seemed to have in common the idea that it was difficult to follow the way parts of the thesis fitted together.

What is most immediately obvious is how low the figures are for the supervisors in the Science and Medical faculties. These findings are consistent with the Stanford University survey (Casanave and Hubbard, 1992) where the Humanities and Social Science staff perceived students to have more problems with development of ideas and overall organisation than did the staff in Science and Technology. One might initially assume from this that science and technology students find logical argument easier than students in the Arts and

Humanities, but it is also possible that the display of logical development in science and technology disciplines does not rely upon language. Logic presents itself through experimental results, formula, figures, statistics etc. Language is not needed as much as in the Arts and Humanities to develop an argument. Again it will be interesting to see if the supervisors' perceptions of difficulties with logical development tally with the HKU students' own perceptions. It would seem from Shaw's student survey that this could well be the case.

The third category of difficulties often referred to has been classified as those relating to STYLE (Fig.3). There was frequent mention of inappropriate style, unacademic style or unscientific style. Quite a wide range exists across faculties here. Poor style was not mentioned at all in Engineering but was seen as quite a problem in the Medical faculty.

Somewhat surprisingly, a number of supervisors ascribed this poor grasp of appropriate style to the fact that the students do not read enough. They feel that the students do not read other dissertations on subjects close to their own nor do they read enough journal articles. However, this suggestion was perhaps more an attempt by these supervisors to find a reason for the poor style they encounter than an actual explanation of the situation as there seems to be no real foundation for their beliefs, particularly as there is evidence that students are aware that reading material of the genre they are expected to produce will improve their style. (See Parkhurst, 1990, and Shaw, 1991, on student strategies for dealing with this).

There seems no reason to believe that the students at the University of Hong Kong are any less aware than students elsewhere of how focused background reading could help them to adopt the desired style but it seems that they are not producing writing that their supervisors find satisfactory from the point of view of style. One possible reason for this could be that it is extremely difficult for students to differentiate styles from one another. Several supervisors considered that this was quite a major problem for L1 speakers so one can fairly safely assume that it must be doubly difficult for L2 speakers. It should be remembered that students tend to be confronted by many examples of informal English, even in an academic setting, for example in lectures and particularly these days in E-mail, so it is not so surprising that they are rather confused about the type of writing that they are expected to produce.

It should be noted here that style was one of the 2 areas of difficulty identified in the faculty of Architecture. The other area identified in this faculty was the fourth of our categories - LEXIS (Fig.3). This was generally seen to be a problem, in much the same way as STYLE, of students using vocabulary that

was too informal and, therefore, inappropriate to the situation, or using a very limited range of vocabulary.

Figure 3: Percentage of supervisors who identified each of four of the problems

	ED.	ARTS	SOC. SCI.	ENGG.	SCI.	MED.
Surface Structures	75	57	83	60	71	86
Logical Development	100	43	50	40	29	14
Style	50	43	17	0	29	43
Lexis	25	29	33	40	0	29

Perhaps the most noticeable point here is the relatively large percentage for Engineering and the fairly small figures in Education and Arts. This might suggest that one feature to investigate is whether the problem with lexis arises from the fact that Engineering has a highly specialised vocabulary with which the students are not sufficiently conversant. However, this possibility seems unlikely given the results of Shaw's (1991) study where, although vocabulary was considered by students to be a frequent source of difficulty, subject technical terms were not prominent. It was mostly semitechnical words and words with multiple meanings which gave rise to difficulties. It could well be that some everyday terms are used in a way peculiar to certain disciplines and the students are not aware of this, as was suggested by one of the supervisors interviewed.

The difference in the figures might also be accounted for quite simply by the fact that the students in the Arts and Education faculties have a wider vocabulary at their command than students in Engineering or Science. This was in fact suggested by a supervisor in Ecology who is in the unusual position of having students coming from first degrees in the Science faculty as well as from the Arts and Social Sciences and is therefore in a better position than most to be able to make a comparison.

Having dealt with the four categories of difficulties that were easily identified, we now turn to the three more problematic categories. These have been categorised as problems of: UNCLEAR MEANING, LACK OF PRECISION

and LACK OF CONCISENESS (Fig.4). Supervisors in Education were particularly troubled about unclear meaning and to a lesser extent about lack of conciseness. Lack of precision was mostly of concern to the Medical Faculty; in fact, it was the second most frequently mentioned problem by supervisors in this faculty.

Figure 4: Percentages of supervisors who noted other problems

	ED.	ARTS	SOC. SCI.	ENGG.	SCI.	MED.
Unclear meaning	100	43	0	40	0	29
Lack of Precision	0	14	0	0	29	57
Lack of Conciseness	50	0	17	0	14	14

The difficulty earlier referred to in categorising these problem areas as distinct items arises from the fact that the supervisors did not identify the source of the problems.

We have used the term UNCLEAR MEANING as an umbrella for expressions such as:

- a) students aren't able to express their intended meaning
- b) there's a lot of ambiguity
- c) they don't seem to be able to get across exactly what they want to say

But without examining individual instances where the meaning is described as "unclear", it is not possible to say whether the lack of clarity arises from inappropriate lexis, incorrect grammatical choice, poor clause structure, misuse of cohesive devices, or, indeed, use of an unsuitable style. Supervisors, in Engineering in particular, complained of students using long complex sentences where they lose control of the sentence and, thus, confuse the reader. If they were to use the simple sentence structures generally recommended in texts on scientific style perhaps students would be able to convey their meanings more clearly.

The lack of conciseness complained of, variously described as 'being verbose' or 'using 10 words where 3 would do', could also be essentially a lexical or stylistic problem. Students may simply lack the vocabulary to be concise or consider that to display the level of sophistication they feel necessary at postgraduate level they should use long and complex sentences.

The same could be said for lack of precision, but there is a strong possibility that there is another dimension to this particular problem; the students may not recognise the need to be precise. This perception that precision is not necessary in a dissertation could arise from what Shaw in his 1991 paper calls the "pseudocommunicative nature of the dissertation task" (p.194). By this, he is referring to the fact that the dissertation writer is well aware that the readers for whom the writing is intended, that is, primarily, the supervisor and examiner, already know a large part of the information that is contained in the writing and, therefore, little real communication is taking place in terms of information transfer. What the writer is actually doing is trying to display knowledge to someone who is already an expert in the field. If it is this expert that the writer has in mind as an audience, then precise explanations may not be considered necessary. This is exactly what a Chinese student in Shaw's study observed about the situation regarding the writing of a dissertation in China. In fact, 64% of the students interviewed in Shaw's survey did actually have such a reader, including supervisor and examiner, in mind when writing. Moreover, as Shaw pointed out, some supervisors may themselves compound the difficulty by advising students to write for an expert or for someone with considerable background knowledge of the situation. If precision is required of dissertation writers, they may be better advised to imagine, as one of Shaw's interviewees actually did, that they are writing to teach the material to an educated and interested lay reader.

(vi) *How difficulties are dealt with*

Having looked at the difficulties that were most frequently perceived by supervisors, we can look at how these difficulties are generally dealt with. Not all the supervisors interviewed agreed that helping students with their language problems along with the content of the dissertation was, or should be, part of the responsibility of a supervisor. Many provided help only because they felt there was no alternative; it was at times necessary to deal with language problems in order to get through to the content, but they would prefer to be able to concentrate on the subject matter .

A total of 23, i.e. 50% of those interviewed so far, thought it was part of their job and 18 disagreed. (There were five who felt unable to come down on one side of the fence or the other on this issue). One important point that should be mentioned here is that this was the only feature in our study so far where there was a marked difference in the attitudes of the L1 and L2 supervisors. Only 28% of the L2 supervisors interviewed, as against 64% of the L1 supervisors, considered help with language to be part of their responsibility as a supervisor. Any attempt at an explanation of this difference at this early stage of investigations would be pure conjecture, but it might well be a point worth bearing in mind for later study.

However, regardless of their feelings on the matter, all supervisors did deal with language problems and, for the most part, what they did in all faculties except one was fairly similar - quite straightforwardly, they corrected mistakes. The stringency and the point in time of the correction varied considerably: some corrected only early drafts, some all drafts, some later drafts, some only the major mistakes, some the first page only, some copiously, some practically rewrote large tracts. But they all corrected.

The only faculty where the supervisors tended to adopt a different approach was Education where supervisors sometimes corrected but frequently suggested changes in sentence structure or organisation, indicated where rewording would be beneficial or recommended peer editing. In other words, they tried to use the feedback sessions as a teaching tool to improve future writing. There seemed to be little awareness on the part of most supervisors that simply correcting all errors may not be the best way to help students in the long term and it would, in fact, be unrealistic to expect such awareness of subject tutors who are not, after all, language specialists.

The supervisors may well, however, be using a method which the students find most pleasing! As Harris and Silva remarked in their 1993 paper on tutoring ESL students, "students often come to the writing centre seeking an editor, someone who will mark and correct their errors and help them fix the paper" (p.530) .

Unfortunately, correcting will not generally create better writers, only a better piece of writing.

(vii) Views on provisions needed

There does seem from the interviews so far conducted to be some measure of dissatisfaction, and in some cases even exasperation, amongst supervisors about the writing that research students produce. There is certainly a large gap - hence the title of this paper, "The Quantum Leap" - between the type of writing that undergraduates are expected to produce, that is in most cases an essay or report, and the type of writing demanded of research students in their dissertations. Many students do not manage to make the leap from one type of writing to the other which may well account for the large number of uncompleted doctoral dissertations throughout the world!

Although some supervisors thought that students needed no additional help with writing in English even though problems were evident initially because they would gradually acquire the necessary writing skills as they progressed in their studies, a large proportion were in favour of students being offered some

kind of assistance by the university in making the leap, above and beyond the kind of help that they as supervisors felt able or willing to offer.

Approximately 20% were in favour of continuing some form of 1:1 tuition on a purely voluntary basis as is presently offered by the Writing Support Service in the English Centre at the University of Hong Kong. At the other end of the spectrum was a small group (9%) who wanted a compulsory course or writing for all research students in their first year, much in the manner of the courses offered on American doctoral programmes. However, most favoured courses or workshops that groups of students could attend according to specific needs on the recommendation of their supervisors or as the result of diagnosed problems.

Conclusion

It would seem from the findings presented in this paper that the assumption that students taken on for research degrees are linguistically able to produce a dissertation without guidance in the writing process is, in many cases, far from correct. Students are seen by their supervisors to have problems not only with grammatical structures (for L2 speakers), but also with presenting a logical argument, using appropriate style and lexis, and conveying meaning in a concise and unambiguous manner. It can be concluded, therefore, that some assistance with writing would be beneficial.

Naturally, not all research students have all of the problems identified by the supervisors who were interviewed. Indeed, individual students will have difficulties with different aspects of writing and it is for this reason that the English Centre at the University of Hong Kong has decided not to run general writing courses for research students but rather to set up workshops where students are expected to work on improving the writing they are producing for their supervisors. The workshops will make use of an analytic framework that has been devised with input from the three sources targeted in our project: the students themselves, their supervisors and samples of their writing. The framework will be used by tutors to diagnose the strengths and weaknesses of a student's writing and by small groups of students to try and help each other identify problem areas. This identification of problems by the students themselves is an important feature of the workshops as it is considered essential for students to learn to diagnose their own problems; after all, the first step in addressing any problem is to recognise its existence.

It is hoped that this approach will enable the students to become aware of the real demands of the dissertation writing task and will, thus, help them to make that quantum leap from essay to thesis.

Acknowledgements

Our thanks go to all the research students' supervisors at the University of Hong Kong who agreed to be interviewed and, therefore, made this project possible. Special thanks go to Ann Prince, our colleague in the English Centre, who conducted many of the interviews, helped to collate the results and gave us valuable comments and suggestions on the early drafts of this paper.

References

- Casanave, C.P. and Hubbard, P. (1992) "The Writing Assignments and Writing Problems of Doctoral Students: Faculty Perceptions, Pedagogical Issues, and Needed Research". *English for Specific Purposes*, 11, 33-49.
- DEET [Department of Education and Training] (1988) " Report on Progress of Postgraduate Award Holders: 1979 cohort." Canberra:APGS
- Dunkerley, D. and Weeks, J. (1994) "Social Science Research Degrees: Completion Times and Rates". In R.G. Burgess (ed) *Postgraduate Education and Training in the Social Sciences*. London: Jessica Kingsley Publishers.
- Hamilton, A. (1990) *Writing Dissertations*. London: RIBA Publication.
- Harris, M. and Silva, T. (1993) "Tutoring ESL Students: Issues and Options". *College Composition and Communication*, 44 (Dec. 1993), 525-537.
- Parkhurst, C. (1990) "The Composition Process of Science Writers". *English for Specific Purposes*, 9, 169-179.
- Phillips, E.M. and Pugh, D.S. (1987) *How to get a PhD. A handbook for students and their supervisors*. Milton Keynes: Open University Press.
- Shaw, P. (1991) "Science Research Students' Composing Processes." *English for Specific Purposes*, 10, 189-207.

- Swales, J. and Feak, C (1994) *Academic Writing for Graduate Students*. Ann Arbor: The University of Michigan Press.
- Torrance, M. and Thomas, G. (1994) "The Development of Writing Skills in Doctoral Research Students." In R.G. Burgess (ed) *Postgraduate Education and Training in the Social Sciences* London: Jessica Kingsley Publishers.
- Van Wageningen, R.K. (1991) *Writing a thesis*. New Jersey: Prentice-Hall, Inc.

14 DIALOGUE JOURNALS AND COURSE EVALUATION

*Jonathan Shaw and Terry Clayton¹
Asian Institute of Technology, Thailand*

Introduction

In March 1992, we arrived in Ho Chi Minh City to teach a two-month intensive pre-session language course¹ to teaching staff from four universities, the first element of an ambitious Swiss-funded program to develop management education in Vietnamese universities. The general learning objectives of our language course were these:

- to become more effective communicators in English in academic contexts
- to develop an awareness of the relevance and importance of their own experience and knowledge to the study of any discipline, not least language
- to develop an understanding of teamwork and collaborative approaches to learning and decision-making;
- to develop the confidence and skills necessary for self-directed learning in general, and for self-directed language learning in particular;
- to start considering seriously questions of learning and teaching methodology, with respect to higher education in general.

These objectives reflect various aspects of what we were trying to do. One of these was to make sure that program participants were ready, from a language point of view, to start a heavy academic program in a field of which few had any direct experience. But because these were university teachers, embarking on a process of change, we also wanted to start to engage them in discussions about how we worked. And in all of our objectives, we envisaged what we were doing as the start of a process, and that participants would continue to develop.

But given these objectives, how were we to evaluate the effectiveness of what we were doing? Even for more conventional language courses, it has long been the case that "... what is observable and measurable might be less important than what is more difficult to observe or measure" (Alderson, 1986). Moreover, particular problems could be anticipated in our attempting to achieve such goals in Vietnam. As Denham (1992) observes, Vietnam is

¹ Correspondence to: Center for Language and Educational Technology, Asian Institute of Technology, GPO Box 2754, Bangkok 10501, Thailand.
E-mail address: clet@emailhost.ait.ac.th

norm-dependent, possibly to a greater extent than other countries in Southeast Asia, and this, we had reason to believe, was especially the case in matters of teaching and learning. Earlier trips to Vietnam had shown us that teaching in Vietnamese universities consisted predominantly of teachers reading aloud from books and writing on blackboards whilst students remained passive. In other words, we anticipated a certain degree of resistance to the innovative, experiential approach which we wanted to offer. However, the overall project goal was to transform education. Participants were aware of this, and recognized the need for change, at least in principle. From the outset, our intention was to provide a deep-end introduction to the fourteen-month graduate program in management, and this is reflected in the objectives.

Several recent articles have addressed the question of the use of logbooks or journals in language courses. Vanett and Jurich (1990), for example, have argued that the collaborative use of journals can transform the traditional roles of teachers and learners in language classrooms. Savage and Whisenand (1993) have proposed that logbooks can reflect how an intensive language workshops' objectives were realized by learners, and thus have potential as a source of qualitative data.

This paper looks at the nature of the teacher-student interactions which emerged in the dialogue journal, with a particular focus on the authenticity of our responses to what students were writing about. The paper further attempts to illustrate the quality of interactions, and argues that these interaction, particularly those which address issues of teaching methodology comprise data which might be used for purposes of course evaluation. Finally, we consider how such journals might be used as instruments for formally evaluating human resource development courses.

Dialogue Journals

Within the literature, various terms are used to describe journals kept by students as one component of a course. We prefer the term dialogue journals, because it reflects the interactive nature of the technology in a way that other terms - student diaries or logbooks - do not. However, these three terms are used more or less interchangeably in the literature by language teachers and researchers who use the technology to address both writing skills and broader humanistic goals. Research into their use has indicated that they are an effective means of improving fluency and the mechanics of writing (Ross, 1988; Peyton, 1990; Reyes 1991). Language teachers at all levels have long recognized the value of dialogue journals in addressing issues of motivation,

self-image, expression and personal empowerment (Spack and Sadow, 1983; Jackson, 1988; Meath-Lang, 1990).

The interest in journals has not been confined to language courses. Winer (1992) examined the use of journals in a teacher-training course in tracing the process by which practice and reflection on specific activities change awareness of and attitudes toward the teaching of writing. Similarly, Jarvis (1992) uses journals as a means by which teachers on an in-service training course might reflect on their learning. Jarvis discovered that the journals also "functioned as feedback about learning, about attitudes to the course and anxieties about the course. I have found the records give me a necessary sense of being in contact with my learners' learning..." (p. 141).

Other practitioners have suggested that journals can be used in processes of course evaluation. Oscarson, (1989) proposes that the day-to-day record of learning experiences in journals make journals good self-assessment instruments. Nunan (1988: 132-4) proposes a more structured approach, with students responding regularly to a series of questions dealing explicitly with aspects of the language course to be assessed.

Savage and Whisenand (1993) used journal entries to describe the realization of objectives in an intensive five-day language workshop in north-east Thailand. They state that the journals their students kept served a dual evaluative purpose: to assess progress in writing for specific purposes, and to evaluate the workshop itself. This work started to explore the possibilities of journals as a source of data relevant to course evaluation. In addition, they sought to develop instruments which could be used for the design of future language learning activities for particular students. It is these perspectives that we intend to explore and develop in this paper.

The Educational Development Context

The goal of the Swiss-funded project is to transform Vietnamese business management education, and the principal means of achieving this was to retrain existing university teachers. The teachers chosen had all received their bachelor degrees in either engineering or economics disciplines. The retraining consisted principally of a general two-year Masters of Business Administration degree, after which participants would return to their universities and work on developing curricula appropriate to the emerging free-market situation in Vietnam.

As university lecturers struggling with the need to make curricula in universities relevant to the rapidly changing economic realities in Vietnam, participants were aware of the need for change. Many were already using Western textbooks on economics, marketing and organization theory in their courses, often translating the textbook for students, replete with examples from America; others had been educated in technological disciplines, had no experience of teaching management, but had nevertheless been given responsibility for developing courses and teaching various management fields. However, as Fullan (1983: 36) makes clear, educational change is multidimensional:

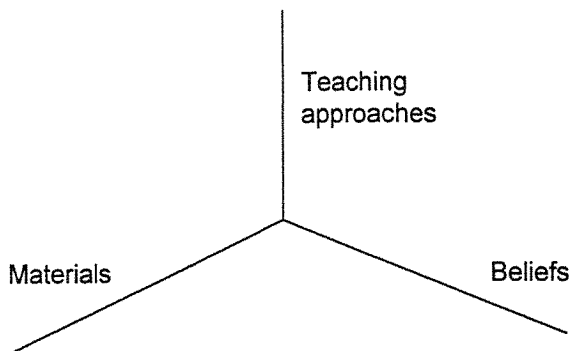


Figure 1: The Multidimensionality of change
from Fullan, 1983:37

Objectively, it is possible to clarify the meaning of educational change by identifying and describing its main dimensions. Ignorance of these dimensions explains a number of interesting phenomena in the field of educational change: for example, why some people accept an innovation they do not understand, why some aspects of a change are implemented, and not others, and why strategies for change neglect certain components. (Fullan, 1983: 37)

For Fullan, the dimensions of change are materials, teaching approaches and beliefs, and change involves a dynamic interrelationship of the three dimensions. Change inevitably causes anxiety, loss and struggle for those affected once they realize that basic conceptions of education and change are involved - "their occupational identity, sense of competence and their self-respect" (1983: 40). Schön (1971) similarly argues that real change involves passing through zones of uncertainty; this uncertainty cannot be pre-empted by planning.

When the course began, our knowledge of the Vietnamese system of higher education was limited. Jon had visited Ho Chi Minh City briefly in the months prior to the project, and met with many of the participants from the south. These trips allowed us to collect enough information about the participants and the system within which they worked to know that the course we wanted to offer would be possible. We did know that, in common with most other educational systems in Asia, it was dominated by a Confucian philosophy of education, dependent on a conduit model of learning and explicit knowledge, and that various physical and financial constraints appeared to, and were believed to, militate against the introduction of other curricular structures and educational paradigms. There were also signals that the kind of language course proposed was not felt to be appropriate in a Vietnamese situation. Faculty of industrial management recommended that we make use of classical poetry if we wanted to motivate participants, implying that they had heard, from a colleague then studying at AIT, that our approach was boring.

The Course

The language course we designed for this program is described in detail elsewhere (Hall and Kenny, 1988; Kenny, 1993; Clayton and Shaw, 1993), and explicitly sought to change more than participants' knowledge of language by challenging assumptions about the nature of knowing and learning. Our transformative agenda was independent of the main management program, partly of necessity, because the curriculum of this program was unwritten at the time. The participants of the Swiss program were university teachers, with little experience of what businesses - particularly those with foreign interests, such as joint ventures - actually did. Getting out of the classroom, we believed, was essential to provide participants with meaningful experience of the world in which international business occurs. In common with practice on our university's pre-sessional language course, Talkbase, we wanted participants to get out of the classroom almost as soon as they first entered it. The first task of going out of the classroom to find out about a word (Hall and Kenny, 1988) would send the immediate signal that this would not be a course in which knowledge flowed from teacher to passive student. Early tasks would develop this theme, focusing on collecting information on business in Ho Chi Minh City (described in Clayton and Shaw, 1993); later investigative research work (Kenny, 1993) would allow participants to build on the experience. Methodologically, our course would probably stand in sharp contrast to anything participants had previously experienced.

We did not consider questions of how we would evaluate the effectiveness of what we were to do before starting the course, and there were no formal assessments built into the curriculum. This omission was partly the result of the haste with which the program began. Two months before we arrived, it was still not clear exactly when we would start. Furthermore, we have for some time been dissatisfied with instruments used at AIT to evaluate the success of courses and programs there. Finally, as is the case in so many development projects, language is tacked on to the "real" program and there is little, if any, demand for evaluation or accountability.

Our Use of Dialogue Journals

Our understanding of the importance of the dialogue journals to the success of the language course emerged as the course evolved. In designing the two-month pre-session course, we included dialogue journals, principally because dialogue journals were used at AIT. However, we probably saw them more as means by which participants could practice writing. What we knew about dialogue journals from the literature supported the view that they helped students to 'improve' their writing skills. Like Jarvis (1992), however, we discovered different purposes as we used them and became progressively more excited by the dialogues being developed, and the possibilities these interactions offered for both evaluative purposes and for dealing with potential conflict. The former perspective is the one examined here.

On the first day of the course, participants were given an empty notebook with the instruction that they were 'to record their ideas and impressions during the eight-week course'. They could write anything they wished; they could, if they chose, simply keep a record of what had happened to them during the day. The rationale we offered was that we believed that writing such a journal would help them to think through their ideas, and that this would improve their writing skills. No specific writing tasks were assigned, and there was no specific compulsion on participants to write anything at all. We told participants that we would collect the journals regularly and respond to whatever they wrote. We wanted them to focus solely on content and to take responsibility for their own accuracy, and explained to participants that we would not be correcting language in the journals. In Peyton's terms (1990:xi), we did not intend to play the default role of evaluator of students' writing¹¹.

The extended dialogue between each participant and ourselves was not the result of a conscious decision on our part, but evolved over the early weeks of the course. Terry and Jon's differing understandings of the journals' purpose

and function within the course are reflected in the differing nature of comments we made in early responses to participants' entries. In the first week, these entries rarely focused on substantive matters and often seem to reflect some confusion about where we were going. Terry's responses were from the beginning friendly, chatty, picking up on possible problems that participants might have been having in settling in to the city, or relating with each other while Jon made comments reflecting his own didactic understanding. This led him both to try to move students away from using the journal as an notebook and to ask questions aimed at further clarifying meaning. Thus:

Dai (after several days of notes, but little reflection): I have a new way to learn English. It is active and effective method. It is good for me, but I have some difficulties. I need some time to get used to learn English. And I need some books, for example, Oxford Dictionary,, business dictionary. I think we need to learn how best to read, how best to write and how best to speak.

Terry Yesterday you were talking to me while Jon and I were waiting for the taxi. I enjoyed it. You are an easy person to talk to, especially after a few 333 (Saigon beer)! So relax. Stop worrying about your English. Don't think about these things, just do it. How about your social life? Meeting any new people? Know any good bars we can go to?ⁱⁱⁱ

Jon (to Hai) (1) What do you mean "give us some justifications"?

(2) You will have [topics to write about.] What kind of topics do you think you need?

(3) No time - really?

Keep writing - you're doing fine. I'd like to know more about what you think, and why: here you describe what has been happening.

Our responses to the first week's entries were generally aimed at encouraging participants to write about their reactions. We both wanted to know what participants were thinking, whether they liked it or hated us for it, whether they were responding as professional educators or as students - in other words, we started to explicitly encourage, although not compel, participants to reflect on what was happening educationally. In this sense, what we were doing might well - in other circumstances - have been for the purpose of formative evaluation.

A typical first week response was

Terry Lan - welcome to the world of computers! Losing files happens under the best of conditions. Now you know why we say SAVE OFTEN.

So you think we are amusing do you!
Most people reported feeling very much like you the first days Relax.
There is such a thing as trying too hard. How do you feel now?
Keep writing. Write more. Where do you live? Have you got any
suggestions for me whilst I'm in HCMC? I want to go places and see
things that tourists don't usually see. (End of week 1)

and typically, this response was longer than Lan's entire entries for the first week.

We soon found that we were spending a lot of time with many of the students socially. They were taking us up on suggestions that they drink beer with us, and were searching out places which might interest us. At the end of the first week, the participants from Hanoi invited us to a small party. To their considerable surprise, we accepted their invitation, and having entrusted ourselves to cyclo taxi drivers, we surprised them again by appearing at Mr. Tam's roomy apartment on Dien Bien Phu Road, two hours late and rain-sodden, to drink beer, and eat Vietnamese ham sandwiches, and to discuss how they really felt about what we had been asking them to do during the first week.

Much of Huyen's first week entries consisted of her trying to settle into Ho Chi Minh City, and her homesickness. She also started to write about some of the problems she and her colleagues were having with the tasks we had set:

The greatest problem of making contact with foreign companies is that I haven't any means of transportation and I don't know the way to go there. I think that we should do it together with the south students - who live in HCM City.
HELP US, PLEASE!

Terry, in addition to consoling her, sought to direct her toward reflection on the experiences of business people coming to the city: His full response to her first week's writings was:

Huyen - Rent seems to be very expensive in HCMC. This always happens when there is a very low vacancy rate. Were you and Mrs. Nguyet friends before [the course] or not? So what is this apartment like? Can you describe it?

Is this the first time you have been away from your daughter? How does your husband feel about you being away from home all the time? Do you think this will help your husband and daughter have a better relationship?

OK, so its difficult contacting foreign businesses. Now you know some of the problems faced by foreign business people when they come to HCM City and they can't even speak Vietnamese.

Keep writing. Great party Friday evening.

When we returned journals to participants after the first weekend, many were disappointed that we had not 'corrected' their 'bad English':

Thuy Till now, I'm also worried about my writing, because when I received my logbook, I didn't see any correction in it. So if I ~~have~~ write sentences in a wrong way, I will do it again.

Jon But actually your writing is highly expressive. I understand your quite subtle nuances. Really!

We knew from the formalized classroom consultations and informal discussions - such as those at Tam's apartment - that many students had doubts about what we were doing. We expected the journals to emerge as a possible - less ephemeral - locus of these discussions. However, few participants were willing to discuss their doubts in writing at the beginning of the course. In view of Vietnam's still dominant centralized political culture, this was not surprising. For many participants, concerns with correction and other aspects of our approach which deviated from the language teaching methodology norm in Vietnam were our entry into a broader consideration of course objectives. Other participants needed less encouragement to start reflecting of the nature of what we doing and its broader applicability:

Trung (a statistics teacher) I'm thinking about your method. I've worked as a teacher for 13 years. I've learned by heart the course I teach. In my classes, I work almost - it makes my students passive, and they feel so tired because they have to write so much. They don't have time to think.. In future, I will apply your method into my work. I think after introducing my students generally what they have to do, I'll divide them in to groups. And I'll ask these groups to prepare in detail their work to represent to class. This will make my students more actively and they can understand the subject clearly. (Beginning of week 2)

As participants became more confident about using the journals, they started to write more about a greater variety of experiences. The following exchange was typical of the middle weeks of the course:

Lan It's long time I haven't went to Ben Thanh market. When I am a student in college, some girlfriends and I have something interesting. We like watching everything on foot, go around some shops, although

we don't have the need to go shopping, especially after hard work in the National Library. We went to release our minds.

This morning I go ~~to~~ alone. I just want to buy a pocket cassette that can record, but I go and watch quite a lot and it is not boring to me. Goods are of great variety with all shapes and sizes ~~both~~ in all the colors. The designing of goods is perhaps studied carefully. They are put at our eye level, specially the favourite articles. Goods seem to be changed continuously to satisfy customer taste. ...

Terry This is called 'retail marketing' and people write books about what you just said in 1 paragraph, but they don't actually say a lot more. Very nicely expressed, Ms. Lan.

We were aware from conversations that many of the participants harbored doubts about what we were trying to do. Such doubts focused on an insecurity engendered by our asking participants to do real things with language, rather than reassuringly integral exercises.

Trung It seems that you overestimate our English. Sometimes you speak with us as with the Americans. In the class, although I always try to listen to you, I still misunderstand you many times.

On Jon's birthday, to my deep regret I cannot come. Because on that day, my uncle who lives in the provinces came to visit me. I have to stay at home to have lunch with him. I'm even absent in the afternoon to take him to some relatives. Very sorry.

Tan^{iv} It's interesting the text analysis sessions have helped you so much with critical reading. More practice, will I am sure, develop your ability to become an efficient reader. . Do Jon and Terry overestimate your English or do you underestimate?

Terry So is your problem getting any clearer? I think so Its a very natural - and a necessary process. A good process. Keep going..

In the second month of the course, participants were focused on project work, in Kenny's words, becoming "investigative researchers" (1983). Comments on language teaching methodology were superseded by concerns with research method and working together in groups.

Trung It's really difficult to persuade others to change their opinions. When I read what Mr. Thuong writes, I don't feel fluently. Theirs are something vague and confused - Lam, Nhan and I tell him that, but finally nothing is changed. He always tries to preserve his own opinion. Each person can have his own way to examine a problem. For instance, in my opinion, this course is about language, computers and methodology. But Jon can have different opinion about it. In this public reportback, you suggest me to present our piece of work. Its

easier for me to present when I arrange the facts in my own way. So you can see some differences between what I say and what you read in the article.

As you say, people like to speak with foreigners when they study language. But sometimes they have nothing to say with foreigners, especially the introverts as I. Moreover, I am a teacher so I am afraid of saying something wrong, as "I'd like dog food" instead of "I'd like dog meat". This weakness makes me slow progress in learning language. So I like to drink beer with you. In a party besides the communication, I can speak more easily "beer in word out".

Terry Trung your comments are so interesting, so insightful, that I really wish you would write more often. Jon's going to comment on some particular points of what you've said, so I will leave these aspects to him. What I want to comment on in particular is your remark about your "mistake" about "dogfood". First of all, how would you ever learn that your particular way of expressing that notion was an error (not a mistake) unless you expressed it? My clarification was a form of correction. Miss Lan had a similar experience She was upset once when I wrote that she was more "talkative" in a small group She thought that "talkative" had a negative connotation which it can have in a particular context. The point is she talked to me about it and so came to deepen her understanding of that particular word and to its usage. I could tell you a dozen stories of errors I have made in Thai. The fact is that I have learned quite a lot of Thai by making errors. Personally, I enjoy the laugh it gets from native Thai speakers and I can get more laughs telling the story to other people later on. Language can be a lot of fun that way. Sometimes I make the error on purpose just for the reaction I know I will get. If you want to accelerate your progress in learning English, allow yourself to make more errors. (May 5)

Jon Trung - I'm not sure if there is a correlation between what we have to say, and the number of words we need to say it. It seems anomalous to me that an 'introvert' should feel compelled to practice - if we do not want to speak our own language, why speak in English? The poverty of practice. In fact, there is a huge amount of evidence to suggest that the kind of practice favored here does not help - it is only useful when meaningful communication takes place. Terry is right about mistakes - we have to make them. But I also see how hard it is when faced with one's own students. Your solution to this is to learn by heart - but your presentation yesterday indicated a flexibility which belied the claim that you had 'learnt by heart'.

Different opinions - what do you think about my opinion that this program is about education, knowledge and power?

Could this program use different techniques to minimize the problem of an old teacher, Trung, standing in front of former students, scared to make mistakes?

As students' project work developed, they did increasingly become investigative researchers, and reflections on what this meant appeared more frequently in the journals:

Thuy Ernst and Young, we received a letter from this company. They refused to answer our questionnaires, Mr. Andy Lee, manager of the office in S. Vietnam (*sic*) says that his company also make a questionnaire with this problem, but they can't let us know the result. Are we disappointed? No, because it's also an answer. It means that the problem we want to find out is a real problem.

Terry Yes! They are real problems and because they are real problems and not made up classroom exercises or role plays, you are learning real things. "Learning by doing" means doing real things.

Course Evaluation and Dialogue Journals

Jon continued working with the program, developing and implementing language support classes and workshops for the program. The dialogue journals were a rich source of information on the program participants to which we were able to refer throughout the remaining 24 months of the program, and from the point of view of evaluation, the deep understanding of what participants were thinking, what concerned them as teachers and problems which they faced outside of the classroom.

The dialogue journal interactions provided a huge quantity of data from which it is possible to determine the effectiveness of what we had been doing. The data is probably most relevant in its raw form, which provides a rich picture of students' responses to a course which we understood to be an educational innovation in Vietnam. Our management colleagues also found the data useful. They did not start arriving until near the end of the pre-session course, and we gave them some of the dialogue journals to read. However, we were surprised and disappointed by some of their responses. All were arriving from universities where very different educational standards prevailed and where different, Western, understandings of knowledge and learning were the norm. So where we saw an impressive level of language accomplishment, learning from experience and reflection on learning experiences, they saw something else - maudlin homesickness, a concern with

insignificances, childishness even. We realized that it takes patience to read through such volumes, particularly if one starts from the first entries and not the latest, and at that time, we were all busy setting up the project. If some way could be found by which the interactions could be systematized, perhaps the value of the raw data could be increased.

The table on the following page is an attempt to relate the dialogue journal interactions to the broader humanistic goals of the course. The basic unit used in the matrix here is an interaction between students and teachers. This might be either ourselves responding to students, or students responding to our comments. For every teacher-student or student-teacher interactional instance, an entry is made in the matrix, with the number corresponding to the week of the course which the interaction occurred. Journals are marked week 1, week 2, week 3, etc. with tabs when we collect them so that the interactions in the chart can be checked against the actual entries.

Weeks 1,2,3,4,5,6,7,8

Name	to develop an awareness of their own experience & knowledge	to develop an understanding of teamwork	to develop the confidence and skills for self-directed learning	considering questions of learning and teaching methodology	Comments
Cuong					
Dai					
Duoc					
Hai					
Huy					
Huyen					
Lam					
Lan					
Nga					
Nguyet					
Nhan					
Oanh					
Phuong					
Phuong					
Quy					
Quyen					
Tam					
Thu					
Thuong					
Tin					
Tri					
Trung					
Vu					

There are, of course, limitations in using such a matrix. The journals almost certainly cannot be taken as wholly representative of the full extent to which the objectives were or were not realized - as discussed above, many interactions took place in discussions in bars, restaurants and class, and we do not know the extent to which certain participants remained cautious of committing anything like criticism to paper. Lam, certainly not a cautious individual, preferred to make his journal entries in Wordperfect on a laptop computer at home, but because he was learning word-processing for the first time his entries were sketchy and very incomplete.

If dialogue journals are to be able to generate data which will be useful for evaluative purposes, then, certain factors need to be present.

- **Authenticity of interaction.** Arnold (1991) has made the point that authenticity is a quality not of materials but of interactions. How can we assure that the interactions in our dialogue journals are authentic? Focusing on genuine concerns of our participants was, we believed, the best way to ensure this. If participants felt that we should have been doing something which we were not doing, this became a matter of genuine concern to them and something which could be negotiated, either in the journals or elsewhere.

Our dialogue journals can be understood to be negotiations between teachers and students. However, this did not mean that the approach was open to modification if participants did not like it. For us, negotiation meant that we would be willing to discuss with them - in any medium which they preferred - the principles on which our approach was based, why, for example, we chose not to focus on form. Probably, few were convinced by our arguments, but a huge quantity of authentic interaction was generated by the difference between their expectations of what teachers, and language teachers specifically, should do, and our own beliefs and intentions. In terms of evaluation, therefore, the journals were not used for formative evaluation purposes; possibly to use them thus might detract from the actual value of the data for summative evaluation purposes.

Moreover, change undoubtedly did happen as a result of the journal interactions. Many of the tasks in the course were developed and implemented ad hoc to the emerging situation.

- **Unrestricted content.** It is important, we feel, that students are free to write about whatever they wish in their journals. This meant that at the beginning of our course, participants used the journal for notetaking, for

drafting various written tasks assigned during the first week, for drafting, in English, letters to faraway loved ones. Our responses at the end of the first week sought to guide participants toward reflection on what they were going through in and outside the classroom; to reveal more of what they thought, what worried them, what they noticed, what interested them. If participants responded to these questions - fine; if not, we responded to what they did write.

- By the same token, students should be free to write nothing at all - compelling students to write seems to us to negate the purpose of dialogue journals. Some participants were certainly reflecting deeply on what we were doing, but chose not to put these reflections in writing, but to talk about them with us. When this happened, we would still urge them to try to write their ideas down, to give them some form of permanence.
- **Total engagement.** We were both responding to every one of the journals once a week (our part-time Vietnamese colleagues would also take a stack and respond as well). This usually meant hunkering down for the evening, and talking thorough our responses together. But in addition to this, we were discussing their concerns in two-on-one consultations; we were encouraging participants to introduce us to Ho Chi Minh City, and we held parties at the house we shared in District 1. We understood all of these various activities to be part of the course. In other words, for a course like this to really be effective, and certainly with this number of students, leaving the classroom, going home and forgetting about the students is not an option if teachers are really taking students' ideas seriously.
- **Team teaching.** As mentioned above, we usually found ourselves responding to journals together and would often discuss what we were reading, and how to respond, and develop tasks for following days through such discussions.

Of these four factors, the hardest to achieve will usually be the third - most of us need to be able to escape from the classroom at some time, and during our two months in Ho Chi Minh City, this was rarely the case. Arguably, this represents a major weakness of the proposed method.

Discussion

Alderson's 1986 paper, 'The nature of the beast', surveyed the state of the art in evaluation in language programs and courses. His arguments as to what

constitutes an effective educational evaluation remain relevant today. He proposes a list of 10 'desiderata' for educational evaluation:

- collaborative ✓
- both quantitative and qualitative ✓
- look at both process and product ✓
- comprehensive ✓
- systematic ✓?
- needs to identify negative findings as well as positive findings ✓
- convince skeptics ?
- feedback into the system, and to participants ✓
- not be an afterthought ✓?
- increase in relevance through practice and experience ✓
- be reflexive - all evaluation should be an experiment in how to do evaluation ✓

An evaluation based on dialogue journals interactions will have most of these characteristics. Persuading participants of our integrity in seeking out and responding to their thoughts is integral to the process described. It is essentially a collaborative process, although we choose not to ask students to evaluate explicitly specific aspects. The data could be analyzed quantitatively as well as qualitatively, or alternatively, could coexist with a formal quantification^v. The data clearly describe the process of the course; it reveals negative features (some participants clearly do not reflect in the way we would have liked); the data clearly helps faculty in the main program to continue designing learning experiences which are relevant to this particular group of learners.

We are certainly guilty of not having thought ahead at the beginning of the course to how we would evaluate it, and our decision to use this data evolved with the course. Having recognized the richness of the data, we certainly did see it as an experiment in how to do evaluation. Questions of adaptability remain open, as does the issue of whether such an approach convinces skeptics. This is partly for this forum to decide!

This paper is an attempt to describe how students in an intensive language and computer pre-session course in Vietnam responded to a wide variety of educational innovations. We were not concerned explicitly with measuring language improvements, although it is apparent from the journal entries over the eight weeks that participants became comfortable with using written English to describe what was happening, to express ideas and concerns and so forth, and that in doing so, they were using syntactically more complex

features of language, and without careful analysis, it is clear that Semke's (1984) results are supported, despite our not explicitly teaching language features at all. However, simply to measure such items would have little relevance given the broader aims of the course. The problems we are faced with in making the data useable are familiar, and we do not suggest here that there is a truly satisfactory alternative to presenting the raw data. However, we do believe that, given the broader humanistic concerns, the approach outlined above is preferable to the various self-assessment procedures outlined by Nunan (1988), Oscarson (1989) and others.

References

- Alderson, J.C. 1986 The nature of the beast, *Proceedings of 1st CULI Conference on Evaluation*, Chulalongkorn Language Institute: Bangkok, Thailand
- Alderson, J.C. and A. Beretta (eds.) 1992 *Evaluating Second Language Education*, Cambridge University Press: Cambridge, UK
- Arnold, E. 1991 Authenticity revisited: How real is real? *English for Specific Purposes*, Vol. 10, 237 - 244.
- Clayton, T. and J. Shaw 1993 Discovering Resources in Ho Chi Minh City, in Savage, W. (Ed.) *Proceedings of the AIT-RELC Conference on Language Programs in Development Projects*, Asian Institute of Technology: Bangkok, Thailand
- Denham, P.A. 1992 English in Vietnam, *World Englishes*, 11/1, 61 - 69
- Fullan, M. with S. Stiegelbauer 1991 *The New Meaning of Educational Change*, Teachers' College Press: New York, NY
- Hall, D. and B. Kenny 1988 An approach to a truly communicative methodology: the AIT pre-session course, *English for Specific Purposes*, 7/1, 19 - 32
- Jackson, W. 1988, Talking through writing, *Language and Education*, 2/1, 1 - 14
- Jarvis, J. 1992 Using Diaries for teacher reflection on in-service courses, *ELT Journal* 6/2, 133 - 143

- Kenny, B. 1993, Investigative Research: How it changes learner status, *TESOL Quarterly*, 27/2, 217 - 231
- Leavitt, H. 1991 Socializing our MBAs: Total Immersion? Managed Cultures? Brainwashing? *California Management Review*, Summer 1991, 127 - 147
- Legutke, M. and H. Thomas 1991 *Process and Experience in the Language Classroom*, Longmans: Harlow, Essex
- Lynch, B. 1992 Evaluating a program inside and out, in Alderson, J.C. and A. Beretta (eds.) 1992 *Evaluating Second Language Education*, Cambridge University Press: Cambridge
- Meath-Lang, B. 1990 The dialogue journal: Reconceiving curriculum and teaching. In J.K. Peyton (ed.) 1990 *Students and teachers writing together: Perspectives on journal writing*, TESOL: Alexandria, VA
- Nunan, D. 1988 *The learner-centred curriculum*, Cambridge University Press: Cambridge
- Oscarson, M. 1989 Self-assessment of language proficiency: rationale and applications, *Language Testing*, Vol. 6 NO. 1 June 1989, 1 - 13
- Peyton, J.K. (ed.) 1990 *Students and teachers writing together: Perspectives on journal writing*, TESOL: Alexandria, VA
- Reyes, M. 1991 A process approach to literacy using dialogue journals and literature logs with second-language learners, *Research in the teaching of English*, Urbana, Illinois
- Ross, S. 1988 First language composition pedagogy in the second language classroom: a reassessment, *RELC Journal*, RELC: Singapore, 19/1, 29 - 48
- Savage, W. and R. Whisenand 1993 Logbooks and language learning objectives in an intensive English for Specific Purposes workshop, *TESOL Quarterly*, 27/4, 741 - 746
- Schön, D. 1987 *Educating the Reflective Practitioner*, Jossey-Bass: San Francisco
- Semke, H.D. 1984 Effects of the Red Pen, *Foreign Language Annals*, 17/3, 195 - 202

- Spack, R. and C. Sadow 1983 Student-teacher working journals in ESL freshman composition, *TESOL Quarterly*, 17/4, 575 - 593
- Vanett, L. and D. Jurich 1990 A context for collaboration: teachers and students writing together. In J.K. Peyton (ed.) 1990 *Students and teachers writing together: Perspectives on journal writing*, TESOL: Alexandria, VA
- Winer, L. 1992 "Spinach to chocolate": Changing awareness and attitudes in ESL writing teachers, *TESOL Quarterly*, 26/1, 57 - 80

Notes

-
- ⁱ In this paper, program refers to the whole management development educational program, course refers to the intensive two-month language and computer pre-session course.
- ⁱⁱ The theoretical justification for not correcting error is provided by Semke (1984), whose research was carried out with American students of German. We assume here that Vietnamese students respond in the same way.
- ⁱⁱⁱ The dialogue journal entries used in this paper are neither complete daily entries nor our complete responses, but excerpts.
- ^{iv} Tan was one of three Vietnamese teachers who worked with us part-time. The times at which Tan was able to work with us meant that he spent much of his time reading and responding to journals, and talking to us about what he was doing and saying.
- ^v we concur fully with Legutke and Thomas's (1992:244) conclusion that summative evaluation of learners' progress against a set of targets remains problematic in task- and project-based curricula, and that any product-oriented view oversimplifies the educational process. We hold this to be as true in adult learning as elsewhere.

X08303065

